

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
12 September 2002 (12.09.2002)

PCT

(10) International Publication Number
WO 02/070539 A2

(51) International Patent Classification: C07K

(21) International Application Number: PCT/US02/05095

(22) International Filing Date: 5 March 2002 (05.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/799,451 5 March 2001 (05.03.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:

US 09/799,451 (CIP)
Filed on 5 March 2001 (05.03.2001)

(71) Applicant (for all designated States except US): HYSEQ, INC. [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). ZHOU, Ping [US/US]; 7595 Newcastle Drive, Cupertino, CA 95014 (US). GOODRICH, Ryle, W. [US/US]; 4896 Sandy Lane, San Jose, CA 95124 (US). ASUNDI, Vinod [US/US]; 709 Foster City Boulevard, Foster City, CA 94404 (US). ZHANG, Jie [CN/US]; 4930 Poplar Terrace, Campbell, CA 95008 (US). ZHAO, Qing, A. [CN/US]; 1556 Koser Road, San Jose, CA 95118 (US). REN, Feiyao [US/US]; 7703 Oak Meadow Court, Cupertino, CA 95014 (US). XUE, Aidong, J. [CN/US]; 1621 South Mary Avenue, Sunnyvale, CA 94087 (US). YANG, Yonghong [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). MA, Yunging [CN/US]; 280 W. California Ave., #206, Sunnyvale, CA 94086 (US). YAMAZAKI, Victoria [JP/US]; 883 Portwalk Place, Redwood Shores, CA 94064 (US). CHEN, Rui-hong [US/US]; 1031 Flying Fish Street, Foster City, CA 94404

(US). WANG, Zhiwei [CN/US]; 836 Alturas Avenue, #B36, Sunnyvale, CA 94085 (US). GHOSH, Malabika [IN/US]; 620 Iris Avenue, Apt. 133, Sunnyvale, CA 94086 (US). WEHRMAN, Tom [US/US]; CCSR Mol Pharm 3210, 269 W. Campus Drive, Stanford, CA 94305 (US). WANG, Jian-Rui [CN/US]; 744 Siendahl Lane, Cupertino, CA 95014 (US). WANG, Dunrui [CN/US]; 12252 Pepper Tree Lane, Poway, CA 92064 (US). DRMANAC, Radoje, T. [US/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo PC, One Financial Center, Boston, MA 02111 (US).

(81) Designated States (national): AE, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

without international search report and to be republished upon receipt of that report

— with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

WO 02/070539 A2

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. BACKGROUND OF THE INVENTION

1.1 TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

1.2 BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

2. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as

allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

5 The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

10 The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1 – 948. The polypeptide sequences are designated SEQ ID NOS: 949-1896. The nucleic acids and polypeptides are provided in the
15 Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

20 The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1 – 948 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1 – 948. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1 – 948 or a degenerate variant or fragment thereof. The identifying
25 sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1 – 948. The sequence information can be a segment of any one of SEQ ID NO: 1 – 948 that uniquely identifies or represents the sequence information of SEQ ID NO: 1 – 948.

30 A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information is provided on a nucleic acid array to detect the polynucleotide that contains the segment. The

array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-948 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-948 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., *Science* 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 949-1896; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1-948; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1-948. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1-948; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in the Sequence Listing.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include

polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-948; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, e.g., *in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the

polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical
5 condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant
10 protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions.
15 The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a
20 method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal
25 antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that
30 modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (e.g., bind to) the polypeptides of the invention. The invention

provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound binds to a polypeptide of the invention is identified.

The methods of the invention also provides methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can effect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 2); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

3. DETAILED DESCRIPTION OF THE INVENTION

3.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only some of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonucleotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and

N is A, C, G or T (U). It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual
5 nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of
10 nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most
15 preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of
20 mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NOs:1-948.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA
25 as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular
30 Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NOs: 1-948. The sequence information can be a segment of any one of SEQ ID NOs: 1-948 that uniquely identifies or

represents the sequence information of that sequence of SEQ ID NO: 1-948. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1+4^{25}$) times the increased probability for mismatch at each nucleotide position (3×25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids,

more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

5 The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full length protein which may include any leader sequence or any processing sequence.

10 The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue.

15 The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer

20 attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant" (or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, e

25 g., recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

30 Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected

in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (e.g., microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins

endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (*e.g.* Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2):134 -143) and factors released from damaged cells (*e.g.* Interleukin-1 Receptor Antagonist, see Arend, W.P. et al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (*i.e.*, hybridization to filter-bound DNA in 0.5 M NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (*i.e.*, washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligos), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" or "substantially similar" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the

corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more than 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, the nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, more preferably at least about 80% sequence identity, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least about 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, *e.g.*, using the Jotun Hein method (Hein, J. (1990) *Methods Enzymol.* 183:626-645). Identity between sequences can also be determined by other methods known in the art, *e.g.* by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

3.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1 – 948; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO: 949-1896; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1 – 948. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO: 1 – 948; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing as SEQ ID NO: 949-1896; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 949-1896. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include all of the coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1 - 948 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1 - 948 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1 - 948 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, more typically at least about 85%, 86%, 87%, 88%, 89%, more typically at least about 90%, 91%, 92%, 93%, 94%, and even more typically at least about 95%, 96%, 97%, 98%, 99% sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1 - 948, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, *e.g.* 15, 17, or 20 nucleotides or more that are selective for (*i.e.* specifically hybridize to any one of the polynucleotides of the invention) are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1 - 948, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NOs: 1 - 948 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology result for the nucleic acids of the present invention, including SEQ ID NOs: 1 - 948, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for Basic Local Alignment Search Tool is used to search for local sequence alignments (Altschul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using Fastxy algorithm.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with

more distant choices (e.g., hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., *supra*, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are

capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention can be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-948, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NOs: 1 - 948 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NOs: 1 - 948 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one

of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsIX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed

recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intra-muscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

3.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1 - 948, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a

double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO: 949-1896 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO: 1 - 948 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO: 1 - 948, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine,

inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α -units, the strands run parallel to each other (Gautier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et*

al. (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

3.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of an mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO: 1 - 948). For example, a derivative of Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a SECX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, SECX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-1124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaire *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to

another molecule, e.g., a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

3.5 HOSTS

5 The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are
10 in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part,
15 the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition
20 to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding
25 sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation
30 (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include

Saccharomyces cerevisiae, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, e.g., inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but

configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

3.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 949-1896 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NOs: 1 - 948 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NOs: 1 - 948 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 949-1896 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 949-1896 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, typically at least about 95%, 96%, 97%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 949-1896.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S.

McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

5 The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature
10 form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed.

15 Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

 The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments
20 which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

 A variety of methodologies known in the art can be utilized to obtain any one of the
25 isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly
30 useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally
5 does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising
10 growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide
15 can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow
20 known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, e.g., Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*.
25 Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well
30 known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well

known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 949-1896.

The protein of the invention may also be expressed as a product of transgenic animals, e.g., as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, e.g., U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the cMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, e.g., Invitrogen, San Diego,

Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

5 The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which
10 will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

15 Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently
20 purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

25 Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

30 The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, *e.g.*, targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability.

Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

3.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), pFam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobicity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

3.7 CHIMERIC AND FUSION PROTEINS

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the invention. In another embodiment, a fusion protein comprises at least two biologically active

portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide according to the invention and the other polypeptide are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus, or to the middle.

5 For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein.

 In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

10 In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a
15 protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin
20 fusion proteins of the invention can be used as immunogens to produce antibodies in a subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

 A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different
25 polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional
30 techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT
PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many

expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

5

3.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

20

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

25

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the

polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional *CAD* gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, e.g., inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element.

Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultschi et al., each of which is incorporated by reference herein in its entirety.

3.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals,

preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

3.10 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

3.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA

sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

3.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the

polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

3.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION

ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin- γ , Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current

Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; 5 Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and 10 Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current 15 Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et 20 al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

3.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent 25 stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotent or pluripotent state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. 30 The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal

cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotent/pluripotent stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotent/pluripotent mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for

inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., *Differentiation*, 48: 173-182, (1991); Klug et al., *J. Clin. Invest.*, 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering* eds. Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. *Proc. Natl. Acad. Sci. U.S.A.*, 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., *Blood*, 77: 2316-2321 (1991).

3.10.5 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines,

thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures

in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

3.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention

contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

5 Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

10 Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

3.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

15 A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or
20 down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including
25 infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

30 Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also be useful in the

treatment of allergic reactions and conditions (e.g., anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxicol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation

may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., *Science* 257:789-792 (1992) and Turka et al., *Proc. Natl. Acad. Sci USA*, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in

cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19;

Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al.,
5 Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production,
10 Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek,
15 D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

20 Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995;
25 Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins
30 that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of

Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad. Sci. USA 88:7548-7551, 1991.

3.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

3.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts,

neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. *J. Clin. Invest.* 95:1370-1376, 1995; Lind et al. *APMIS* 103:140-146, 1995; Muller et al. *Eur. J. Immunol.* 25:1744-1748; Gruber et al. *J. of Immunol.* 152:5860-5867, 1994; Johnston et al. *J. of Immunol.* 153:1762-1768, 1994.

3.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostasis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of

thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those
5 described in: Linet et al., *J. Clin. Pharmacol.* 26:131-140, 1986; Burdick et al., *Thrombosis Res.* 45:413-419, 1987; Humphrey et al., *Fibrinolysis* 5:71-79 (1991); Schaub, *Prostaglandins* 35:467-474, 1988.

3.10.11 CANCER DIAGNOSIS AND THERAPY

10 Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing
15 malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation,
20 inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies
25 including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal
30 neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system,

bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Kaposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention (including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine IICl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguanzone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wiley-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cell lines are available, e.g. from American Type Tissue Culture Collection catalogs.

3.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14. Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited to, ricin.

3.10.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product
5 libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science* 282:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide
10 and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol.*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996)
15 (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then
20 tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The
25 toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

3.10.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a
30 ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used to identify polynucleotides encoding binding partners. As another example,

affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.*, increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications *i.e.* phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

3.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation

associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflammation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic myelogenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

3.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

3.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

- (i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;
- (ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;
- (iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;
- (iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;
- (v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;
- (vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;
- (vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and
- (viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.
- Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:
- (i) increased survival time of neurons in culture;
- (ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, e.g., choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction *in vivo*.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, e.g., weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

3.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition

(including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

3.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect

the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

3.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holositz, et al., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

3.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

3.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01 µg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1 µg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

3.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2,

G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF- α and TGF- β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hyl, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, e.g., treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

3.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in

the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

5

3.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition

for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dye-stuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene

glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable

polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia,

trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μ g to about 100 mg

(preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize

a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF- α and TGF- β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either *in vivo* or *ex vivo* into cells for expression in a

mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

3.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate *in vitro* assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50} . Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of

administration and dosage can be chosen by the individual physician in view of the patient's condition. See, e.g., Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 $\mu\text{g/kg}$ to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 $\mu\text{g/kg}$ to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

3.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

3.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and

immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} , F_{ab}' and $F_{(ab)2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule
5 obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

10 An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An
15 antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 949-1896, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid
20 residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of alpha-2-macroglobulin-like protein that is located on the
25 surface of the protein, e.g., a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art,
30 including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, e.g., Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more

domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full-length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells,

neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

3.13.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g.,

aluminum hydroxide), surface-active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant
5 (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific
10 antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

3.13.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique
20 heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen-binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse,
25 hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or
30 a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-

103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused,
5 immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high
10 level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for
15 the production of human monoclonal antibodies (Kozbor, *J. Immunol.*, 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding
20 specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107:220 (1980). Preferably,
25 antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640
30 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

3.13.3 HUMANIZED ANTIBODIES

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeven et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539). In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise

substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

3.13.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al. (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals that are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the XenomouseTM as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

5

3.13.5 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab)2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab)2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

3.13.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion

preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab')₂ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')₂ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')₂ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was

able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (scFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

3.13.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimide and those disclosed, for example, in U.S. Patent No. 4,676,980.

3.13.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

3.13.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain,

alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies.

5 Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimide HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as
10 glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-
15 methyl-diethylene triamine-pentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionuclide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation
20 using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

3.14 COMPUTER READABLE SEQUENCES

25 In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical
30 storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer

readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NOs: 1 - 948 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NOs: 1 - 948 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means

having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access
5 manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are
10 used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA
15 (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the
20 less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif.
25 There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include,
30 but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

3.15 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 15241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

3.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound

antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

3.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

3.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NOs: 1 - 948, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

- (a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and
- (b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the

complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time
5 sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the
10 activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of
15 the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling
20 techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed"
25 when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or
30 pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or

rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

3.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NOs: 1 - 948. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from any of the nucleotide sequences SEQ ID NOs: 1 - 948 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used

in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

5 Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide
10 sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include *in situ* hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or
15 flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent *in situ* hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical
20 chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject
25 invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

3.20 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced
30 using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be

achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6) 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all references being specifically incorporated herein.

5 Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8) 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any
10 surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

 Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed CovaLink NH. CovaLink NH is a polystyrene surface grafted with
15 secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal. Biochem. 198(1) 138-42).

 The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has
20 been described (Rasmussen *et al.*, (1991). In this technology, a phosphoramidate bond is employed (Chu *et al.*, (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an
25 oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

 More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M
30 1-methylimidazole, pH 7.0 (1-Melm₇), is then added to a final concentration of 10 mM 1-Melm₇. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

 Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-Melm₇, is made fresh and 25 ul added per well. The strips are incubated for 5 hours

at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) *Science* 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) *Nucleic Acids Res.* 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) *Anal. Biochem.* 169(1) 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) *PNAS USA* 91(11) 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

3.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.*

(1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods.

5 Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

10 Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24) 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

15 One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *CviJI*, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

20 The restriction endonuclease *CviJI* normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*CviJI***), yield a quasi-random distribution of DNA fragments from the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *CviJI*** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *CviJI*** restricts pyGCPy and PuGCPy, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

30 As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

3.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm², depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm² and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and

methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

4.0 EXAMPLES

4.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

4.2 EXAMPLE 2

Novel Nucleic Acids

The novel nucleic acids of the present invention of the invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 119, gb pri

119, and UniGene version 119) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 121, gb pri 121, UniGene version 121, Genpept release 121). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1- 948.

Table 1 shows the various tissue sources of SEQ ID NO: 1-1896.

The nearest neighbor results for polypeptides encoded by SEQ ID NO: 1-948 (i.e. SEQ ID NO: 949-1896) were obtained by a BLASTP (version 2.0a1 19MP-WashU) search against Genpept, Geneseq and SwissProt databases using BLAST algorithm. The nearest neighbor result showed the closest homologue with functional annotation for SEQ ID NO: 1-948. The translated amino acid sequences for which the nucleic acid sequence encodes are shown in the Sequence Listing. The homologues with identifiable functions for SEQ ID NO: 1-948 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), polypeptides encoded by SEQ ID NO: 1-948 (i.e. SEQ ID NO: 949-1896) were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the Pfam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) polypeptides encoded by SEQ ID NO: 1-948 (i.e. SEQ ID NO: 949-1896) were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the product of all the e-value of similar domains found, the pFam score for the identified domain within

the sequence, number of similar domains found, and the position of the domain in the SEQ ID NO: being interrogated.

The GeneAtlas™ software package (Molecular Simulations Inc. (MSI), San Diego, CA) was used to predict the three-dimensional structure models for the polypeptides encoded by SEQ ID NO: 1-948 (i.e. SEQ ID NO: 949-1896). Models were generated by (1) PSI-BLAST which is a multiple alignment sequence profile-based searching developed by Altschul et al., (Nucl. Acids res. 25, 3389-3408 (1997)), (2) High Throughput Modeling (HTM) (Molecular Simulations Inc. (MSI) San Diego, CA) which is an automated sequence and structure searching procedure (<http://www.msi.com/>), and (3) SeqFold™ which is a fold recognition method described by Fischer and Eisenberg (J. Mol. Biol. 209, 779-791 (1998)). This analysis was carried out, in part, by comparing the polypeptides of the invention with the known NMR (nuclear magnetic resonance) and x-ray crystal three-dimensional structures as templates. Table 5 shows, "PDB ID", the Protein DataBase (PDB) identifier given to template structure; "Chain ID", identifier of the subcomponent of the PDB template structure; "Compound Information", information of the PDB template structure and/or its subcomponents; "PDB Function Annotation" gives function of the PDB template as annotated by the PDB files (<http://www.rcsb.org/PDB/>); start and end amino acid position of the protein sequence aligned; PSI-BLAST score, the verify score, the SeqFold score, and the Potential(s) of Mean Force (PMF). The verify score produced by GeneAtlas™ software (MSI), is based on Dr. Eisenberg's Profile-3D threading program developed in Dr. David Eisenberg's laboratory (US patent no. 5,436,850 and Luthy, Bowie, and Eisenberg, Nature, 356:83-85 (1992)) and a publication by R. Sanchez and A. Sali, Proc. Natl. Acad. Sci. USA, 95:12502-13597. The verify score produced by GeneAtlas™ normalizes the verify score for proteins with different lengths so that a unified cutoff can be used to select good models as follows:

$$\text{Verify score (normalized)} = (\text{raw score} - \frac{1}{2} \text{ high score}) / (\frac{1}{2} \text{ high score})$$

The PMF score, produced by GeneAtlas™ software (MSI), is a composite scoring function that depends in part on the compactness of the model, sequence identity in the alignment used to build the model, pairwise and surface mean force potential (MFP). As given in Table 5, a verify score between 0 to 1.0, with 1 being the best, represents a good model. Similarly, a PMF score between 0 to 1.0, with 1 being the best, represents a good model. A SeqFold™ score of more than 50 is considered significant. A good model may

also be determined by one of skill in the art based on all the information in Table 5 taken in totality.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determined from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al., as reference, were obtained for the polypeptide sequences. Table 6 shows the position of the last amino acid of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

Table 7 correlates each of SEQ ID NO: 1-948 to a specific chromosomal location.

Table 8 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-948, novel polypeptide sequences SEQ ID NO: 949-1896, and their corresponding priority nucleotide sequences in the priority application USSN 09/799,451, herein incorporated by reference in its entirety.

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
Null	mix tissues library	CTL016	52 137 189-192 316-325 529 591
Null	enriched library	CTL021	65 84 169 189-192 311 316-325 406 676 727 782 850
Null	mix tissues library	CTL028	65
Null	PCR products cloning	PCR2V1	34 51 134 189 191-192 224 232 260 311 388 606 623 806
Null	mix tissues library	SUP002	51 96 103 163 216 272 294 311 316-325 328 378 383 388 446-448 450 453 474 481 500 516 610 774 780 885 904 922
adipocytes	Stratagene	ADP001	2 43 51 73 76 88 97 142 166 181 186 188 208 257 262-263 267-270 282 311 316-325 383 386 427-429 459 463 465 493 507 514 522 545 552 572 643 651 667 700 721 740 754 758 778 795 872 881 883 888 947
adrenal gland	Clontech	ADR002	3-6 10-11 13 16 20-21 24 27-28 33 38 48-49 51 53-54 58 66-67 75 88 97 99 124-125 130 140 157-158 179 188 197-198 200 212-214 216 218 224 229 231 237 257 267 279 281-282 288 302 311 326 362 376-377 381 383 396 398-403 429 443 453-454 456 459-460 474 489 515 526 531-

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			532 540 545 550 559 564 568 577 581 586 589 599 605 610 613 631-632 643 648 651 667 670 672 681 684 699 703 706 708 717-718 734 736 751 779 785-786 795 813 817 837 871 876 887- 888 897 904 907 916 921 924-926 948
adult brain	BioChain	ABR012	140 208 311 748 810
adult brain	BioChain	ABR013	51 245 311 316-325 436 717 810 936
adult brain	Clontech	ABR001	12 51 87 142 169 178 180 245 263 286 288 290 295 304 308 311-313 375 379-380 403 425 428 431 458 486 499 503 512 557-558 567-568 606 610 641 651 695 704 730 741 754 766 810 822 827 841 850 864 871 884 897 917 920 925-927 934 946
adult brain	Clontech	ABR006	2 14-15 22-23 29 32-33 49 66-68 83 99 111-112 115 129 131 142 147 153 157 163 169 189-192 200 205 207 212-214 218 221 229 234 256-257 263 272 276 279 282 292-299 301 311 315 340- 343 349 376-377 383-386 388 403 405 407 410 425 438 453-454 460 463 469 474 489 495 499- 500 511 522 531-532 539 541 545-546 551 556 563 565 571 579-583 591 594 606 626 628 631- 632 643 647 651 678 684 691-692 700 717 721 726 730 732 741 744 754 757 769 772 774 782 788 793 810 820 827-828 853 867 869 875 879 897 913 921-922 925-926 933-934 939-941 947
adult brain	Clontech	ABR008	1-2 9-10 13 16-18 23 27-28 30-32 37 39 42-43 46 49-51 66 70 76 80 83 86-87 95-97 109 111- 112 116-117 124 130-131 133-134 136-137 141- 142 146-147 152-157 160 162 169 171 179 184 189-192 195 200-201 206 211-212 216-218 239 247-248 250 252 254-258 261-263 271-272 276 278 282 288 293-295 297 300 302 307 309 311 314-326 328 333 337-341 343 347 349 351-354 358 360-361 367 374 376-378 381 384 388-390 393 395-396 400-403 405 407 409 411 414 418- 420 422 427-429 433 438 440-441 445-447 450 453-455 458 460-461 463 466-470 474 476 486 491-493 496 498-500 507 511 514 520-521 525 527-529 531-532 534-535 542 546 548-549 551- 552 557-558 560 562 564-566 568 571-572 578- 583 586-587 590-591 594 599 602 606 618-619 621 626 629 631-634 643-644 647 651 656-660 664 670 672 677 680 684 687-688 691-695 697 706 709-710 712-714 716-718 721-722 724-725 727-728 730 733 740-741 745 751-752 754 761 765 774 777-779 787 790 792-793 799 801-804 808 810 812 820 822 824 827 831-832 834 836 845 850 858-861 868-869 871-872 875-876 883 887 891 897 900 904 907 910 913 917-920 925- 927 929 931-934 958-941 946-947
adult brain	Clontech	ABR011	51 133 810 892
adult brain	GIBCO	AB3001	16-17 19 40 66 92-94 97 124 131 134 163 186 188 208 213 231 268-270 284 288 295 297 299 311 315-325 340 373 387 396 407 429 469 489 495 498-499 533 542 545 562 568 587 589 618- 619 643 664 687-688 694-695 730 748 836 876 882 884 902 925-926 948
adult brain	GIBCO	ABD003	2 22-24 29 33 43 45 50-51 66 71 75 77 82 87-88

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			91-92 95 131 140 157 179 188-192 200 208 213 220 225 247 252 257 261 263-265 277 284 288 295 299 301 315-325 355-356 373 387-389 392 395-396 407 423 431 443-444 450-451 457 459 468 476 489 495 499-500 514 520-522 532-533 541-542 545-546 557-558 562 564 576-577 581- 583 588-589 591 595 597 599 601 610 619 631- 632 639 643-644 654-655 658-660 664 667 676 682 687-688 693 696 700 704 711 713-714 746 758 765-766 774-775 780 800 802 804 807 810 827 829 834 842 850 854-855 866 870-871 878 892-893 897 899 910 916 920-921 929 931-932 934
adult brain	Invitrogen	ABR014	2 51 65 84 86 134 311 316-325 384 422 445 460 503 525 564 634 651 721 794 804 810 922
adult brain	Invitrogen	ABR015	37 134 263 272 277 294 311 443 467 500 514 582-583 619 651 694 850 871-872 883 888 936
adult brain	Invitrogen	ABR016	19 22 57 134 188 233 271 277 299 373 440 444 459 469 514 640 717 882 890 920
adult brain	Invitrogen	ABT004	1-2 18 28 51 55-57 67 87-88 115 119 137-139 142 163 200 204 213 218 257 263 271 282 288 299 301 311 341 358 370 378 402 407 422-423 427 458 460 463 499 504 534-535 551 557-558 571 586 605-606 610 618 627-628 640 643 680 687 691-692 697 701-702 715 719-721 725 727 753-754 758 771 782 810 827 859 871-872 881 913 920 925-926 938-941 944 946
adult heart	GIBCO	AHR001	1-2 5-6 14-18 20-21 23 28 32 37 41 45 51 53 55- 56 62 66 69-70 80-81 85 87 91 97 107 120-121 124 134 140-141 156 163 165-166 172 188-192 195 197-198 200 208 213 216 221 229 231 235 261-265 267 271 276 284 288 302 305 308 311 316-325 328 333-334 337-338 347 368-369 373 376-377 379-380 389 396 420 440 445 453-454 459-460 465 468 478 483-484 489 491-493 495 501 504 507 514 524 529 533 539 541-543 545 549 552-553 564 566 568 574 577 581-583 587 589-591 596 599 602 605 608-609 618-619 623 625 629-632 643 645 647 651 664 672 676 678 683-684 707 714 716-717 732 735 740 743-744 751 754 757 765 775 778 784-786 788 807-808 810 826 828-829 842 850 860 876 878-880 890 894 897 899 902 916 923-927 933 939-941
adult kidney	GIBCO	AKD001	1-2 5-6 13 16-17 19-23 26 28 33 38-39 43 45 48- 51 55-57 60 66-67 69-73 79 82-83 87 90 94 96- 97 100 103 126 131 134 140 148-149 157 163 166 179 184 186 188-192 200-203 213-216 220- 221 224 226-229 232 235 245 252 257 261-263 268-270 272-274 276-277 279 282 288 290 294 299 308 311 316-325 332 335 339-340 358 360- 363 373 375 379-380 386 388-389 392 395-396 402 413 421 423-424 428-429 431 436 440 444 450 454 457 459-460 468-469 476 489 492-493 499 504 511 513-514 520-521 524-526 531 533 538-542 544-547 552 564 567-568 574 577-578 582-583 590-591 595-596 598 602 607 610 613 618-619 622 631-632 639-642 644 647 651 654- 655 658-659 664 667-669 673 678 680-682 684

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			687 689 693 696 706-707 712 714-715 717-718 721 729-731 734-736 740 744 748 754 760 771 774 782 784 789 795 807 809-810 819 825 834 836-837 842 850 859 870 872 876 878-879 884 887 890 895 897-899 902 905 910 919-921 925-926 933 936 944
adult kidney	Invitrogen	AKT002	1 14-15 28 30 35-37 53-54 73 88 112 114 129 134 137 140 149 157 166 172 186-188 191-192 203-204 213 235 245 257 262-263 266 268-270 273-274 288-289 297 299 302 310-313 315-325 335 340 358 373 378-381 395 413 423 441 450 453 456 459-460 470 477 491-494 500 513 540 542 545 554 556 564 567 587 591 619 622 627 633 643 668-669 677 684 689 693 701-702 704 714 729-730 754 758 760 777 781 785-786 788-789 807 836-837 840 849-850 872 876 881 890 895 905-906 913 923 925-926 931-933 944
adult liver	Clontech	ALV003	159 179 189-192 201 219 257 349 392 568 664 753 796 887 934
adult liver	Invitrogen	ALV002	5-6 28 35-36 52 54 70 72 86-87 103 112 127 134 140 159 179 188 200-201 213 218-219 225 239-240 257 263 271 275 311 315 367 373 388 392 444 459-460 464 468 497-499 512 527 532 542 545 562 599 605 629 640 657 680 684 687-688 706 713 715 717-718 721 742 754 758 771 791-793 818 829 843 854-855 871 878-879 887 921 933-934
adult lung	GIBCO	ALG001	5-6 16 28 38 51 74 97 122 124 134 140 163 188-192 200 218 221 262-263 268-272 294 311 316-325 379-380 429 463 468 493 511 520-522 537-538 542 545 568-569 595 622 643-644 664 667 711 714 721 730 754 775 850 860 863 879 887 897 925-926 944
adult lung	Invitrogen	LGT002	2 5-8 13 16-17 29-31 35-39 43 46 57 67 72 76 78 81 85 87 90 94 97 100 110 119 122 130-131 134 137 140 146 149 167 172 174 179 188 197-198 201 213 216 218 220-221 223 231 245-246 251-252 256-257 262-263 267-270 277 284 288 296 299 301-302 311 316-325 340 354 373 379-380 388 392 395 400-401 410 413 421 431 436 441-443 445 451 455 457 460 463-464 467 469 475 478 489 491 493 497 499 504 507 514 518-519 524 529 534-535 537 542 545-546 548 552 553 559 568 578 581-583 592 597 602-603 605-607 613 615 619 621-622 636-637 642-643 646-647 654-655 679-681 684 687-689 693 701-702 704 706 711 713 715-716 718 727 732-734 738 748 753-754 757-758 760 762 766 769 774 782 785-786 802 817 829 834 850 853 859-860 866-867 870-871 878-879 887 890 899 902 904 910 917 923 925-926 936-937
adult spleen	Clontech	SPLc01	33 38 57 67 75 87 134 142 163 216 221 229 244 257 304 307 311 316-325 340 353-356 378 441 468 525 538 545 560 564 599 721 754 766 780 794 827 841 850 866
adult spleen	GIBCO	ASP001	2 14-15 20-22 29 38 43 48 51 53-56 65 67 72 74 84 87 131-132 134 137 140 172 188-192 200 212 221 256 263 271 282 308 311 316-325 343

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			383 389 423 436 441 443 459-460 467 469 495 499-500 505 514 520-522 524 529 537 539 545 552 585 619 631-632 639 643 664 673 707 723 735 742-744 758 771 799 810 817 836 850 878 925-926 934 936
bladder	Invitrogen	BLD001	5-6 8 20-21 28 72 91 122 126 130 166 188 197 200 213-214 225 257 262 315-325 341 409 486 491 572 593 622 650 673 691-692 810 813 861 870 877 883 887 904
bone marrow	Clontech	BMD007	65 76 84 245 516
bone marrow	Clontech	BMD001	8 13-16 28 38 43 45-48 50-51 57 62-63 65 67 84-85 97 100 104 118 122-124 131 134 140 163 188 214 216 221 224 231 245 252 261-263 268- 270 273-274 279 288 290 311 373 378 389-391 395 414 428 431 436 440-441 443 451 455 459- 460 465 469-470 475 495 497-498 502 507 514- 519 529 537-538 542 546 550 552 556 560-561 563-564 568 576-577 580 587 589 596 601-602 610-613 619-620 626 642-643 647 651 664 666 668 676 678 681-682 684 696 704 706-707 715 727 730 732-735 740 748 753 758 761 764 771 775 780 794 800-801 830 834 836 842 850 863 871-872 878-879 882 884 888 897 900-901 904 910 921 923 929 934 947
bone marrow	Clontech	BMD004	65
bone marrow	GF	BMD002	1-2 5-6 10 13 16-21 27 31 38 42-43 46 57 65-66 76 80 84 87 97 99 110 112 118 131 134 137 140 145 161 163 165 172 195 206 208 221 229 231 237 244 247 252 256 267-270 272 276 278-279 282 284 288 294 301 304 307 311 316-327 333- 334 337-338 345-347 352 360-361 368 373 376- 378 381 383 388 414 436 441 443 450 452 454- 455 457 469-470 483-484 486 490 498 516 519- 521 524 530-531 539 542-543 545-546 551 553 555 559 564 571 576-577 580 585 591 594 602 604-605 607-608 610-612 619-621 625-626 629 631-632 639-640 644 650-651 664-665 684 687- 688 693 699 703 714 723-724 727 733 735 740 742 745 748 750-752 754-755 777-780 784 787 794-795 802 809 817 824 827 831-832 834 846- 847 850-851 854-855 861 867 875 878 883 886 891 894 897 900 902 910 914 919 921 925-926 929 936 939-941 944
cervix	BioChain	CVX001	3-4 14-16 20-23 25 33 42-43 45 48-50 54 57 67 69 75 85 87 91 95-97 107 110 114 124 126-127 131 134 137 140 150 157 163 165 172 185-188 200 204 212-213 216 225 229 245 252 257 261- 263 266-270 276 282 288 290 301-302 308 316- 325 327 340 363-364 372-373 378 383 388-392 394 396 409 413-414 421 428-429 438-440 443- 444 454 456-457 459 463 467 475 486 489 493 495 507 514-515 522 534-537 556 568 572 574 577 582-583 587 594 600 608 610 613 622 626 633 639 643 647-648 651 653 667 680 683 685- 686 693 696 703-704 706 711 721 723-725 727 730-731 734-735 742-743 748 754-757 762 771 776 785-786 788 794 800 802 807 809-810 817 827 829 834-835 842 850 857 860 862-863 868

Tissue Origin	Library/RNA Source	HTSEQ Library Name	SEQ ID NOS:
colon	Invitrogen	CLN001	870 873 876-877 879-880 884 887 891 897 904-905 910 916-917 921 925-926 933 937 947 19-21 53 55-57 72 88 133-134 168 213 245 252 311-313 316-325 340 443 459 469 483-484 486 497 515-516 597 606 622 643 667 676 706 718 742-743 753 766 829 833 872 887 902 923 929 305 311
diaphragm	BioChain	DIA002	
endothelial cells	Stratagene	EDT001	1-2 7-8 14-16 19-22 24 28-29 32-33 41 43 45 51 57 61 74 83 87-88 97 103 112 116-117 131 134 137 140 148 165 172 179 188-192 197-198 208 212-213 220-221 225 229 231 237 246 252 256-258 261-265 268-272 276-277 279 281-282 284 286 288 294 297 299 302 307-308 311-313 326 334-335 340 355-356 358 360-361 364 375 383 386 389 392 403 413 423-424 429 440 443 445 451 453 455-456 459-460 462-463 465-466 468-470 475 491 495 497-499 504 514 520-522 524-526 528 532-536 539-540 546 551-552 554 556 564 566-567 571 574-577 581-583 587 591-593 597-599 601 607 615 618 622 625 633 639 641-644 651 667 677 680 684 691-692 701-702 704 716-717 720-721 726 732-733 735 743-744 754 758 765 785-786 795 802 806 809 819 826 828-830 832 834 836 846-847 850 867 871 877-878 880-891 897 902 907 921 923 925-926 944 946 188
esophagus	BioChain	ESO002	
fetal brain	Clontech	FBR001	33 49 51 126 134 197-198 264-265 360-361 413 460 647 810 819 871
fetal brain	Clontech	FBR004	137 156 205 282 284 405 424 480 489 701-702 820 921
fetal brain	Clontech	FBR006	2 9-10 18-19 22 28 30-32 37 39-40 42-43 46-47 49 57 66-67 76 80 83 96 109 112 116-117 120 124 131 133-134 136 142-143 146 152 155 160 162 165 169 173 184 189-198 200-201 205 215-216 238 244 248 254-255 257-258 260-263 272-274 276-277 282 288 293-294 307 309 311 314-328 343 347 351-352 354 357-358 360-361 373-375 378-381 390 392 400-401 403 405 407 410-411 413 420 424 429 445 450 452-453 458 460 463 467-469 472 474 477 479 483-484 491 499 507 520-521 525 527 529 531 533 538 545 551 562 564 566 571 574 579 581-583 587 591 599-600 604 606 611 626 629 631-632 638 643 651 654-655 657-660 672-673 676-677 684 689 693-694 697 699 709 714-715 717 720-721 732-733 735 744 748 751-752 754 761 763 767 772 775 777-779 781 785-786 790 792 802 804 808 810 820 824 826 838-840 850 858-860 864 866 872-873 881 891-892 901-902 904 910-911 913 917-918 920 925-926 933 939-941 946-947 316-325 684
fetal brain	Clontech	FBR003	
fetal brain	GIBCO	HFB001	2 12 16-17 19 23 27-28 32-33 39 41-45 49 87-89 94 97 100 107 112 130-131 134 142 157 163 172 188-192 200 216 224-225 231 237 242 246 252 258 261 263-265 271 273-274 276-277 288 295 299 301 307 311 314-326 328 341 355-356 373 375 387 389 392 395 424-425 431 438 445 450-452 457 459-460 468-469 475 489 491 495

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			500 504 511 514 520-529 531 533 540-542 545 552 554 557-558 566 576-577 579-584 587 591 596 598-599 606 613 626 631-632 643 651 664 668 673 676 680 693-694 696 703-704 716-717 721 727 735 738 740 744 748 757-758 769 774 778 780-781 810 827-828 830 850 869 871-872 876 878-879 884 890 892 897 899 904 906-907 913 916 918 920 924 928 934 938 946
fetal brain	Invitrogen	FBT002	2-4 20-21 45 51 53 57 88 93 125-126 134 166 184 186 188 200 213 224 263 276-278 307 311 341 373 375 418-419 423 427 432 450 452 459- 460 470 492 498-499 507 514 522 534-535 545 550-552 571 577 610 714 721 743 754 795 827 861 866 872-873 887 896 925-926 934 939-941 946 948
fetal heart	Invitrogen	FHR001	2-4 10 13 16-18 29 31-32 37-38 43 46 49 51 53 55-56 67-68 75 80 85 87 97 115 120 137 152 156 160-161 163 168-169 174 178 189-192 196 200 216 220 225 252 262 276-277 282 288 301- 302 305 311 315-325 333 343 351 357-358 360- 361 368-372 378 424 436 440-441 445 453 460 469 478 483-484 495 520-521 527 533 538 541- 543 546 556 564-566 568 576 581 587 594-595 601-602 606 609 612 615 633 638 640 643 653- 655 664-665 672-673 677 684 691-693 697 704 707 709 717 735 738 744 746 748-749 751-752 754 761 777 779 781-782 785-786 797-798 820 824 826 829 834 838 841-847 850 875 877-878 893-894 897 901 910 913 925-927 936 946
fetal kidney	Clontech	FKD001	8 14-15 32 43 50 68 96 106 126 131 134 140 186 188 226-228 233 279 282 311 339 428 440 450 456 468 552 618 651 700 726 735 748 751 781 794 797-798 826 878 887 899
fetal kidney	Clontech	FKD002	50 83 96 131 134 143 163 172 193-194 201 203 215 263 273-274 311 316-325 339 360-363 374 376-377 379-380 388 394 400-401 403 407 425 440 451 454 493 525 536-538 540-542 572 580 582-583 587 605-606 621 631-632 647 673 689 706 709 714 726 735 761 774 777 799 809 845- 848 858 872 875 878-879 882 895 918 927
fetal kidney	Invitrogen	FKD007	66 214
fetal liver	Clontech	FLV002	52 189-192 219 297 308 335 364 378 427 828
fetal liver	Clontech	FLV004	2 19 28-29 37 39 49 52-53 55-56 62 65-66 76 87 124 134 137 139 142 179 188 195 208 216 219 244 252 263 268-270 272 277 287-288 294 303 305 311 315-325 339 355-356 358 360-361 368 374 378 403 441 454-455 460 477 483-484 497 514 520-521 542 553 582-583 587 591 594 611- 613 620-621 638-639 654-655 658-659 681 684 687-688 709 721 730 738 744 752 754 781 793 802 813 818 826 832 836 854-855 876 878 893 897 900 910 924 933 944
fetal liver	Invitrogen	FLV001	2 7 19 28 35-37 47 52 54-56 66 95 134 139 179 188-192 200 213 218 263 272 288 294 305 311 315 349 378-381 388 392 403 426 443 454 459- 460 469 496-499 514 527 529 532 534-535 553 586 605 640 644 658-659 673 680 687 698 713 715 720-721 723 726 754 758 778 795 817-818

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
fetal liver-spleen	Soares	FLS001	829 853 861 868 870-871 897 903 933 1-23 25-39 41-43 45-46 49-65 76 83-84 87-88 91 94 97 100 112 118 122 126 129-130 134 137 140 148 163 168 172 179 186 188-189 191-192 197-198 200-201 213-214 216 221 225-229 231- 232 235 242 252-253 256-258 262-271 277 284 287-288 297 299 307-309 311 315-326 330 355- 356 360-362 370 373 378-380 388 392 394 396 400-402 413 426 428 436 440-441 443-447 450 454-455 457 459-460 463-465 467-469 475 477- 478 489-509 511-514 519-521 525-527 529-535 537 539 542 545 551-552 555-556 559 561 567 569 571 576-577 580-583 586-587 589 591-592 595 598-602 605 607 610 612-613 618-619 623 625-626 631-632 638-640 642-644 646-647 649 651-652 654-655 667 673 676 680-681 683-684 700 703-704 706 711 715-718 720-721 726 732- 735 740 742-744 748 754 756-758 763-765 771 774-775 777-782 785-786 790 793-795 797-798 806-808 815 818 824-826 829-830 834-835 837 841 846-847 849-850 856 860-861 866 870-872 876 878 881-883 888 894 897-898 902 905 907 910 919 924-926 929 933 942 947
fetal liver-spleen	Soares	FLS002	1 3-4 11-12 14-17 20-23 26-29 32-34 38 41 43 45-47 49 51-52 55-62 65-67 76 83-85 87-88 90- 91 95 97-99 104-105 112 114-117 126 130 133 150 163 165 172 178 186-187 193-194 200-202 208 213 221 225 229 232-235 244-246 248-253 256-257 262-265 267-271 273-274 284 287 299 311 315 326 335 337-338 343 355-356 358 375- 378 381 392 394 400-402 414 416 426 428-429 440-442 444-447 453-455 457 459 461 464-465 467 476-477 483-484 489-490 492 495 497-500 504 506-507 509 511-514 519 522 524 526 532- 535 537 539-540 542 545 551-552 556 567 569 574 576-577 581 589-590 592 599 601-602 605 607 610 612-613 619-620 625 627 629 631-632 638 640-641 646 648-649 654-655 667 670 683- 684 687-688 693 696 700 703-706 713 716-721 726 734-735 740 742 744 748 754 758 771 775 777-778 780-782 785-787 790-792 794-799 801- 802 806 808 818 824 829 835 849-850 852-855 857 870-871 876 882-884 886 888 890 894 897- 898 900 902-903 907 919 921 923-926 929-930 933 938-942
fetal liver-spleen	Soares	FLS003	30 34 67 85 88 99 130 172 188-189 191-192 213 229 231 257 311 315-325 329 331 335 362 391 394 400-402 423 441 455 457 461 476 498 500 511 523 531-532 537 542 576 587 592 612-613 625 649 665-666 703 719 731 733 740 744 771 775 777-778 787 797-798 819 824 826 850 854- 856 861 863 870 879 884 897 923 931-933 947
fetal lung	Clontech	FLG001	65 179 213 223 340 360-361 491 564 577 591 627 646 650 712 715 744 758 939-941
fetal lung	Clontech	FLG004	388 921
fetal lung	Invitrogen	FLG003	49 54 75 97 137 148 152 188 197-199 213-214 225 240 256 288 316-325 369 378 392 423 429 464 496 526 580-581 586 591 693 706 726-727

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			766 878 913 925-926 939-941
fetal muscle	Invitrogen	FMS001	28 65 115 121 126 134 137 156 168 172-173 181 213-214 225 263 267 305 340-341 360-361 440 459 516 534-535 543 564 586 606 609-610 623 650 676 683 754 766 853 871 886 894 930 934-935 948
fetal muscle	Invitrogen	FMS002	19-21 41 49 51 53 57 75 96 101 103 112 134 136 156 171 184 188 191-192 212 216 250 262 267 276 305 311 342 348-350 355-356 360-361 374 392 403 411 415 423 425 457 469 491 495 499 508 515 517 534-536 543 546 564 566 576 580 582-583 587 594 599 609 611 615 618 623 644 647 658-659 664 668-669 677 683 691-692 696 703 735 743 754 766 788 802 817 826 828 850 877-879 894 910 925-926 935
fetal skin	Invitrogen	FSK001	3-7 18 24 27 29 35-37 51 53 55-56 66-67 76 90 95 97 122 126 134 136-137 166-167 181 188- 192 208 213-214 224-225 245 250 252 257 260 262 268-271 273-274 282 284 297 302 312-313 315-326 341 367 373 375 378 383 387-388 390 394 423 429 440-441 450 454-455 457 459 463- 464 470 472 475 486 489 492 495 498-500 511 514 524 527 530-532 534-535 541-542 545 550 552 555-556 565 571 586 589 591-592 602 604 606 610 613 618 622-623 626 631-632 640 651 654-655 672-673 685-686 693 701-702 704 706 717-718 720 723 727-728 744 754 762 764-766 768-771 795 809 814 821 824 827 843 853 868 870-874 887-888 890 897 902 907 925-926 928 930 933-934 939-941 944-945 947
fetal skin	Invitrogen	FSK002	2 5-6 19 29 34 51 57 59 88 97 101 124 131 134 143 163 166 172 189 191-192 196 212 216 222- 223 231 250 257 263 268-272 282 284 287-288 294 297 299 302 304 310-311 316-325 328 333 340 352 360-361 365-367 372 379-380 388 390 400-401 403 410-411 440 449-450 454 457 463 470 478 491 495 500 505 515 520-521 524 532 534-535 541 555 560 562 564 572 576 581 592 595 599 611 622 626 630 636 640 642 650 664 677 683 691-693 696 699 701-702 708-709 715 721 723 728 735 744 747-748 750 754 766 779 782 799 803 807 813 820 824 826 834 846-847 867-868 872-874 878-879 890-891 897 901 904 907 910 912 916 918 925-926 933 944-945
fetal spleen	BioChain	FSP001	311 748
infant brain	Soares	IB2002	2-4 12 14-15 20-21 23-24 27 29 31-32 39 41 46 48-49 51 53 55-56 66 75-76 86-88 93 95 101 105 108-109 116-117 125 127 129 131 136 145 163 166 170 180-181 186 188-189 191-192 200- 201 207-208 212-214 216 220 224 229 231 245 247 252 257 259 264-265 267 271 279 282 288 293 295 299-300 311 314-326 337-338 340 349 367 373 375 388 390-393 396 402 405 407 418- 421 424 428-429 431 433 436 450 452-453 457 459 463 468 489 495 498-500 507 511 522-524 526 528-530 532 541-542 545-546 552 557-558 562 564-566 571 577-583 587 589-591 599-601 606 608 613-614 619 631-632 647 654-655 658-

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			659 667 676 684 691-693 696 700 704 711 718 721 723 725 740-741 743-744 748 754 775 777- 778 780 788 792-793 795 802 805 808 819 826- 829 834 836 838 861 863 869-870 875 879 881- 882 884 887 890-891 893 897 902 920-921 925- 926 934 938-941 946
infant brain	Soares	IB2003	2 27 37 39 43 48-49 51 53 85-87 97 106 113 124 126-127 131 142 166 170 188 200-201 208 214- 215 220 224 226-228 231 251 257 263 267 271- 272 279 288 293 299 311 314-326 337-339 349 360-361 367 386 392 397 400-402 407 410 418- 419 424-425 427 429 452 454 460 475 489 495 497-500 507 522-523 525 529 532 539 542 545- 546 551-552 557-558 564-565 578 582-583 585 591 601 606 625-626 631-633 643-644 673 690- 693 701-702 706 711 721 723 734 740-741 743- 744 748 751 754 761 778 788 795 802 808 819 826-827 829 837 843 869-871 875 878-880 884 896-897 902 920 933-934 946
leukocytes	Clontech	LUC003	12 14-15 18 32 111 134 137 172 221 277 280 311 316-325 436 454 467 549 552 568 585 603 643 691-692 698-699 734 744 751 754 784 797- 798 861 897 916 923
leukocytes	GiBCO	LUC001	2-4 7-8 13-17 20-23 31-33 38 43 48-49 51 53-57 63 66-68 74-78 85-88 93 97 122 124 129 131- 132 134 137 140 163 166 168 171-172 175 188- 192 197 200 208-213 216 221 223 231 236 242 252 257-258 261-263 268-270 272 277 279 287- 288 294 307 311 314 316-326 329 337-339 341 373-374 376-377 381 388-392 396 400-401 413- 414 423 436 441 450 454-455 459 463 465 467 489 491-493 495 498-499 504-505 507 514 518 520-522 524 526 529 531 533 536-537 539-540 545 552-554 556 568 571 577 580 585 589-590 596 599 602 605 607 610 612-613 615 618-619 621-622 625 638 640 642-644 664 667 677-678 684 690-693 696 700 703-704 707 713 715-718 721 727 734-735 738 740-746 748 753-754 758 775 778 780 789 794 797-798 801-802 815 817 825 827 829 834 836 846-847 850 859-861 863- 864 866-867 871-872 878 884 886-888 891 896- 897 902 904 910 913 916 921 923-926 929-932 936 943
lung	Stratagene	LF001	2-4 22 28 32-33 47 51 79 120 129 134 140 163 172 188 208 220-221 231 252 257 263 276-277 284 307 375 378-380 396 423 428 440 450 459 463 486 491 493 495 499 539 571 591 601 607 613 615 618 625 639 651 684 716-717 721 727 735 748 782 828 850 870-871
lymph node	Clontech	ALN001	43 98 131 140 163 188 221 245 277 299 311 491 515 546 564 593 603 610 615 630 682 694 707 717 800 831 850 878 880 936 939-941 947
lymphocyte	ATCC	LPC001	2 16 19-21 25 31-32 49 53 55-56 63 67 85-87 90 97 120 122 137 140 163 165 168 172 188 197- 198 215-216 221 229 231 236-238 248 252 256 272-274 283-284 288 294 299 316-326 343 368 374 378 395 423 431 454-455 467 469 476 478 491 495 498 505 512 515 517-518 520-522 524

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			526 529 531-532 537 539 542 545-546 551 556 571 577 580 589-590 592 596 601 622 631-632 640 642 654-655 664 666 668-669 673 684 703 708 716 718 721 723 727 733 735 743-744 754 758 765 771 775 777-779 783-786 797-799 810 816-817 828 834 845 859 861 863 870 878-879 881 884 887 890 897 904 907 910 912-913 918 923 929 939-942 945-947
macrophage	Invitrogen	IIMP001	49 97 208 252 301 306 311 316-325 337-338 345-346 416 512 522 572 670 716 743 785-786 802 888 919 923
mammary gland	Invitrogen	MMG001	1-8 14-18 20-21 25 28-29 37 39 43 49 51-57 60 66-67 72 75-76 87 95 97 103-104 106 112 115 119 122 127 130 134 137 139 142 150 166-168 172 175 184 186 188-189 191-192 200 213-214 222-224 226-229 240 252 257-259 263 267 271 276 278 282 287-288 299 301-302 305 307-308 311-313 316-325 327-328 332 340-341 358 360- 362 369 373 378 381 383 388 390-392 397 402- 403 409 415-416 423 425 428-429 433 436 444 454-456 459-460 464 467 469 481 483-484 486 493 495 498-499 515 524-525 529-530 532-537 541-542 545 551-552 562 582-583 586-587 593 599-600 602 604-605 610 618-619 622 625-627 634 644 646-647 652 654-655 662 673 676 680 684 687-688 691-692 701-703 715 717 721 723 726 735 743 751 754 758 765-766 771 777-778 789 803 805 807 809-811 821 827 829 850 860 887-888 892 896 898 901-902 905 911 913 917 925-926 930 936 939-942
melanoma	Clontech	MEL004	3-4 16 20-21 43 46 48 97 103 147 163 188 191 213 216 221 231 241 245-246 260 262-263 316- 325 381 407 431 504 525 527 542 556 568 577 589 596 607 613 676 693 714 735 737-739 744 758-760 775 822 850 863 878 887 897
neuron	Stratagene	NTD001	2 16 32 51 66 88 97 124 130 134 137 172 188- 189 191-192 231 252 257 260 277 291 373 424 431 454 460 489 495 523 525 582-583 591 631- 632 643 649 670 695 725-726 735 765 789 797- 798 837 850 878 884 888 890 913 929 946
neuron	Stratagene	NTR001	2 5-6 20-22 136-137 188-194 197-198 224 311 375 381 410 457 462 475 495 531 546 548 552 599 618 678 743 752 819 828 890 895 897 930 934 938 944 946
neuronal cells	Stratagene	NTU001	2 5-6 20-21 55-56 87 137 188-192 197-198 215- 216 260 287 291 310-311 316-325 365 375 423 457 459 470 499 532 542 564 576 598-599 623 643 651-652 673 721 726 743 745 752 754 765 780 787 789 822 829 870 875 888 896 917 919 929
ovary	Invitrogen	AOV001	2-7 10 13 18-22 25 27-28 30-31 33 38-39 41-43 45 48 50-51 53-56 62 66-67 69 72 74-75 80 83 85-87 93 95 99-101 107 112-115 120 124-126 129 131 134-137 140 142-143 147-148 162-163 172 178 188 191-192 200 204 208 212-213 220- 221 225 229 231 235 237 246-247 252-253 258 261-262 264-265 267-272 276-279 282 287-288 290 293-294 299 307-308 311 316-327 332 337-

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			338 340-342 349 360-362 373-374 379-381 386 388-389 393 396 399 403 413-414 423 425 427 429 431 441 444-445 450 452 454-455 457 459- 460 462 467-470 475 477 483-484 489 491-493 495-496 500 504-505 507 515 518-519 522-523 527-529 531 533 537-540 542 545-546 548 551- 552 555 564 568 570-571 577 579 581-583 589- 596 599-600 605 610 613 615-616 619 623 625- 627 630-636 639 641-644 647 649 651 654-655 664 668-669 672 676 678 680-682 684 687 694 701-704 706-707 715-717 721-722 727 729 731- 732 734-735 738 740 743 748-749 753-754 758 764 771 775 777-782 784 800 802 807 821-822 824 828 834 836-837 842 846-847 850 860-863 866 870-871 876-880 882 884 887 890-891 897 899 901 906 910 913 920-921 923-924 933 939- 941 944 947
pituitary gland	Clontech	PIT004	41-42 83 85 97 134 193-194 204 208 213 224 257-258 263-265 285-286 308 311 360-361 413 443 445 491 514 529 532 639 644 647 682 701- 702 716 781 822 829 836 850 933 939-941 947
placenta	Clontech	PLA003	16 31 34 49 66 80 87 97 101-102 134 158 165 172 179 184 188 197-198 209-210 218 220 229 235 249 256 267-270 277 287-288 302 307 332 360-361 365 388 394 414 441 444 454 457 460 493 498-500 505-506 509 529 531-532 550 559- 560 564 572 587 601 625 630-632 638 672 682- 684 689 706 708 726 733 735 744 754 761 784- 786 793 863 875 897 924 929 937
placenta	Invitrogen	APL001	34 68 102 263 444 493 520-521 534-535 689 706 754 797-798
placenta	Invitrogen	APL002	2 14-15 43 55-56 66-67 134 184 213 221 229 252 257 263 277 287 394 443 529 532 618 622 684 742 754 810 829 883 902
prostate	Clontech	PRT001	7-8 51 85 87 97 100 122 134 139 214 216 221 231 257 271 276 335 337-338 392 400-401 431 440 459 477 530 534-535 546 556 582-583 599 622 631-632 639 651 663-664 673 683 707 715 735 740 765 773-774 777 810 823 897 909 919 934 939-941 947
rectum	Invitrogen	REC001	18 54 66 134 137 169 188 200 213 225 251 265 288 311-313 316-325 340 388 423 429 441 454 459 514 532 542 610 626 646 651 657 715 719 723 728 735 740 758 766 785-786 823 829 833 836 886 942
saliva gland	Clontech	SALS03	460
salivary gland	Clontech	SAL001	31 49 78 95 134 136-137 143 176 188 208 223 244 268-270 284 308 311 316-325 388-389 391 436 441 459 476 514-515 520-521 532 545 568 589 596 610 619 684 691-692 713 718 727 736 754 777 824 836 864 867 878 883 897 901-902 916-917 933 938-941
skeletal muscle	Clontech	SKM001	42 98 156 163 191-192 200 261 305 311 395 415 462 468 504 531 543 566 582-583 585 594 680 740 853 875 927 933 935
skeletal muscle	Clontech	SKM002	850
skin fibroblast	ATCC	SFB001	379-380 850
skin fibroblast	ATCC	SFB002	742 850

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
skin fibroblast	ATCC	SFB003	87
small intestine	Clontech	SIN001	27-29 31 38 40 46 48 51 54 57-58 62 65 67 75 77 85 97 110 112 116-117 119 131-132 134 137 140 161 163 166 168 177 188 197-198 208 213 220 224 229 246 257 261-262 264-265 276-277 288 295 297 299 311 316-326 328-330 337-338 340 360-361 373 375 382 390-391 410 413 428- 429 436 438 440 453-454 459 468 476-477 497 507 511 522 531 536 538 542 545-546 548 552 556 564 570-571 576 580-581 586-587 591 596 599 605 610 613 619 625-626 643-644 651-653 664-666 668-670 677 680 684 693 700-702 706- 707 713-715 723-724 729-730 735 740 746 748 753-754 757-758 764 777-778 784-786 818 822 824 826-829 833-837 842 862-863 865-867 877- 878 886 897 900-902 906 913 916 921 925-926 936 939-941
spinal cord	Clontech	SPC001	18 23 33 37 42 51 67 87 92 94 97 100 140 162 184 188 191-192 208 213 220 231 248 262 268- 271 273-274 282 287-288 290 307 311 316-325 358 364 376-377 383 387 389-390 402 412 422 444 455 476 483-484 489 504 522 534-535 556 562 587-588 591 597 603-604 618-619 643 651 667-670 677 693 703-704 717-718 727 746 757 773 808 810 827 834-835 837 850 871-872 875 904 910 931-932 939-941
stomach	Clontech	STO001	18 65 88 163 188 208 213 261 272 277 286 294 336 373 396 412 459 514 553 602 610 647 651- 652 671 673 714 774 790 831 833 842 850 876
testis	GIBCO	ATS001	1 3-4 14-16 28 31 45-46 66 85 90 95 97 103 112 128-130 134 140 163 166 188 191-192 199-200 213 226-228 261-265 267-271 284 302 311 316- 325 327 379-380 391 413 421 428 444 454 457 459-460 467 491 493 495 500 505 519 525 529 532 534-535 545 552 556 566 568 575 596 599 613 616-617 647 649 651 680 684 703 707 716 719 721 727 734 738 740 744 748 758 765-766 774 777 782 802 810 817 827-828 834 842 846- 847 850 862-863 871-872 878 880 892 901 916- 917 921
thalamus	Clontech	THA002	2 87 96 103 106 189-192 208 252 258 295 308 311 367 376-377 383-384 445 455 459-460 498 529 587 598 602 629 654-655 705-706 715 717 723 754 775 810 817 822 864 867 881 892 927 930
thymus	Clontech	THM001	3-4 8 18 28 54 57 63 65 68 84 97 100 116-117 122 134 142 151 169 171-172 188 195 197-198 201 213 221 237 245 261 287 311 316-325 360- 361 376-377 423 441 444 459 489 491-493 495 498 504 507 514 527 532 534-536 539 553 556 568 571-572 590 595-596 599 610 618 622 631- 632 643 647 651 654-655 664 687-688 691-693 703 715 721 733-735 748 760 762 765 781 794 799 802 831 834 836 842 850 860-861 863 871 878 885 896-897 903 910 923 925-926 928 939- 941
thymus	Clontech	THMc02	2-4 17 20-22 37-38 42-43 46 63 65-68 76 88 95 103 118 120 124 134 137 140-141 143 163 165

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			171 179 182 189-194 198 200 212-215 221 226-228 231 244 257 262 266 276-277 287-288 297 299 307 316-325 341 352 358 360-361 373 376-377 379-381 389 391 394-396 403 410-411 436 440 445 450 459 463-464 469 478 491 495 500 507 511 519-521 530 532 539 542 550 555 560 563 576 581 587 595 601 610-611 613-614 618 622 625-626 631-632 638 642-644 657 664 667 670 673 680 683 687 691-693 699 715-716 721 740 743-744 747-754 761 763-765 771 777 780-781 784-787 790 794 805 811 820 826 831 834 841 845 861 867-868 878 881 883 891 893-894 896-897 902 904 910 912-914 918 923 936-941 946-947
thyroid gland	Clontech	THR001	1-2 18-21 27 32 38 42 46 49 51 53-56 66 72 77-78 87-88 97 115 119 124 130-131 134 136 152 163-165 172 183 188-192 202 212-213 216 221 224 229 235 241-243 252 257-258 261 263-265 267 277 279 297 301 305 308 311 316-325 327 357 363 373 376-377 381 383 389 397 400-401 410 413-414 427-428 443-444 446-447 457 459 463 467-469 475 482 489 495 499-500 504 509 513 519-522 526 529 533 537-538 542 545-546 548 556 564 567-568 582-583 589 592 599 605 608 611 621 623 630 642-644 648 651-652 654-655 664 672-676 684-686 691-694 700 706-708 713 717-718 721-722 725 729 731 734-735 740 748 753-754 760 764 766 771 774 777 781 792 797-800 802 805 826 828-829 834 842 850 861 863 868 876 879 897 899 901 910 913 929 937 939-941
trachea	Clontech	TRC001	20-21 38 112 161 163 188 263 267 327 413 420 457 459-460 471 514 540-541 552 572 574 622 639 634-655 676-677 691-692 707 725 743 748 765 777-778 862 868 897 905 908 944
umbilical cord	BioChain	FUC001	1-2 29 32 46 67 83 87 94 134 136 140 148 160 163 166 172 181 186-192 197-198 208 213 216 225-231 237 252 261-265 267-270 279 282 288 295 302 308 311 316-326 339-340 365 376-377 379-380 384 392-397 421 423 428 433 440 445 452 459 461 463-464 470 472 489 491 495 497 500 507 517-518 522 525-526 528 534-535 540 545-546 556-558 564 566 568 571-572 577 592 599 601 605 610 618 623 644 651 661 668-669 673 678 680 685-686 696 706 709 718 735-736 748 754 769 772-777 782 792 797-799 802 807 809 815 817 824 850 854-855 870 876 881 888 891 897 899 901 913 921 928 930-932
uterus	Clontech	UTR001	51 67 126 130 133 140 188-192 229 267 329 373 440 491 514 599 685-686 693 713 716-717 735 897 905 911 939-941
young liver	GIBCO	ALV001	3-4 17 20-21 32 43 55-56 70 100 134 137 163 172 174 179 186 188-192 200 213 216 219 221 229 232 252 275 301 311 315-325 378 381 392 441 459-460 497 499-500 514 524 526 533 539 550 568 571 588-589 595 619 622 631-632 642 658-659 664 677 680 693 700 707 713 719 743 754 757-758 766 807 834 863 867 876 884 887

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			904 907

TABLE 2

SEQ ID	Hit ID	Species	Description	S score	Percent identity
949	AAM25384	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:899.	644	99
949	AAAY27581	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 15.	644	99
949	gil3785618	Mus musculus	sideroflexin 4	396	60
950	gil276912	Homo sapiens	Human putative ubiquitin C-terminal hydrolase (UHX1) mRNA, complete cds.	3719	100
950	gil2653165	Homo sapiens	ubiquitin specific protease 11, clone MGC:8620 IMAGE:2961383, mRNA, complete cds.	3709	99
950	gil3529494	Mus musculus	Similar to ubiquitin specific protease 11	3167	83
951	AAAY11696	Homo sapiens	MITU LckSH3 domain-combining protein.	4131	99
951	AAG78648	Homo sapiens	SHAN- Human RalBP1 related protein 82.	3875	99
951	gil3625166	Homo sapiens	RALBP1 mRNA, complete cds.	3875	99
952	AAAY11696	Homo sapiens	MITU LckSH3 domain-combining protein.	3953	96
952	AAG78648	Homo sapiens	SHAN- Human RalBP1 related protein 82.	3697	96
952	gil3625166	Homo sapiens	RALBP1 mRNA, complete cds.	3697	96
953	gil0437191	Homo sapiens	cDNA: FLJ21146 fis, clone CAS09305.	2190	94
953	gil2843222	Mus musculus	putative	1672	77
953	gil5778893	Homo sapiens	Similar to hypothetical protein FLJ20967, clone MGC:11140 IMAGE:3837082, mRNA, complete cds.	1328	99
954	gil0437191	Homo sapiens	cDNA: FLJ21146 fis, clone CAS09305.	2359	100
954	gil2843222	Mus musculus	putative	1643	72
954	gil5778893	Homo sapiens	Similar to hypothetical protein FLJ20967, clone MGC:11140 IMAGE:3837082, mRNA, complete cds.	1409	99
955	gil5825377	Mus musculus	NIMA-related kinase 8	2009	89
955	gil5825379	Danio rerio	NIMA-related kinase 8	1439	70
955	AAO019	Homo sapiens	HYSE- Human polypeptide SEQ	548	85

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	74		ID NO 15866.		
956	AAW88660	Homo sapiens	HUMA- Secreted protein encoded by gene 127 clone HSUBW09.	175	97
956	AAO00187	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 14079.	70	55
956	gi13815429	Sulfolobus solfataricus	Sugar transport related protein	70	40
957	AAB93966	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14027.	1133	100
957	gi10433835	Homo sapiens	cDNA FLJ12377 fis, clone MAMMA1002524, weakly similar to HYPOTHETICAL 117.8 KD PROTEIN IN STE2-FRS2 INTERGENIC REGION.	1133	100
957	AAO04381	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 18273.	594	100
958	AAB95297	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17525.	885	100
958	gi10434941	Homo sapiens	cDNA FLJ13087 fis, clone NT2RP3002099.	885	100
958	gi16740566	Homo sapiens	Similar to hypothetical protein FLJ13087, clone MGC:15009 IMAGE:3536735, mRNA, complete cds.	807	95
959	AAY27676	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 110.	474	100
959	gi5295951	Oryza sativa	Similar to Herpesvirus papio BRRF2 homolog gene, partial cds.(U23857)	69	41
960	AAG89262	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 382.	352	98
960	AAY30721	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	352	98
960	AAB23615	Homo sapiens	ALPH- Human secreted protein SEQ ID NO: 30.	343	97
961	AAY72605	Homo sapiens	INCY- Human Electron Transfer Protein, ETRN-3.	579	100
961	AAO11627	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 23519.	579	100
961	AAG03941	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8022.	570	98
962	gi14603455	Homo sapiens	ubiquitin-conjugating enzyme E2A (RAD6 homolog), clone MGC:20175 IMAGE:3051041, mRNA, complete cds.	599	79
962	gi4883773	Gallus gallus	ubiquitin-conjugating enzyme	599	79
962	gi14485244	Mus musculus	ubiquitin-conjugating enzyme HR6A	599	79
963	gi14603455	Homo sapiens	ubiquitin-conjugating enzyme E2A (RAD6 homolog), clone MGC:20175 IMAGE:3051041, mRNA, complete cds.	699	90
963	gi4883773	Gallus gallus	ubiquitin-conjugating enzyme	699	90
963	gi14485244	Mus musculus	ubiquitin-conjugating enzyme	699	90

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	44		HR6A		
964	gi16877066	Homo sapiens	clone MGC:24447 IMAGE:4077762, mRNA, complete cds.	362	100
964	gi16877059	Homo sapiens	clone MGC:24437 IMAGE:4075637, mRNA, complete cds.	362	100
964	AAAY94959	Homo sapiens	GEMY Human secreted protein clone mc300_1 protein sequence SEQ ID NO:124.	204	97
965	AAB92993	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11723.	2879	97
965	AAG81364	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:246.	2879	97
965	gi14042380	Homo sapiens	cDNA FLJ14690 fis, clone NT2RP2005270.	2879	97
966	AAB95769	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18703.	2841	99
966	gi10436607	Homo sapiens	cDNA FLJ14207 fis, clone NT2RP3003185, weakly similar to TROPOMYOSIN I, FUSION PROTEIN 33.	2841	99
966	gi12833193	Mus musculus	putative	2375	85
967	AAM25413	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:928.	799	100
967	AAW67863	Homo sapiens	HUMA- Human secreted protein encoded by gene 57 clone HFEBF41.	551	98
967	gi13543811	Mus musculus	Unknown (protein for IMAGE:3591061)	95	33
968	gi10437960	Homo sapiens	cDNA: FLJ21792 fis, clone HEP00441.	5865	99
968	AAE06186	Homo sapiens	HUMA- Human gene 58 encoded secreted protein fragment, SEQ ID NO:248.	3088	98
968	AAE06098	Homo sapiens	HUMA- Human gene 58 encoded secreted protein HSLCX03, SEQ ID NO:160.	3088	98
969	gi12698079	Homo sapiens	mRNA for KIAA1767 protein, partial cds.	4441	98
969	AAM25578	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1093.	3898	98
969	AAE06186	Homo sapiens	HUMA- Human gene 58 encoded secreted protein fragment, SEQ ID NO:248.	3464	98
970	AAAY48359	Homo sapiens	META- Human prostate cancer-associated protein 56.	403	98
970	gi15215966	Homo sapiens	DI.8Q12 gene for hypothetical protein, exons 1-2.	92	53
970	AAR99256	Homo sapiens	UYAR- Natural killer lytic associated protein.	75	37
971	gi6561827	Mus musculus	Klf21a	5684	76
971	gi6561829	Mus musculus	Klf21b	4944	60
971	gi126979	Homo sapiens	mRNA for KIAA1708 protein.	4656	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	61		partial cds.		
972	AAW75079	Homo sapiens	HUMA- Human secreted protein encoded by gene 23 clone HBMCT32.	148	100
973	AAY35921	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 170.	548	99
973	AAM25386	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:901.	494	96
973	AAAY35923	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 172.	494	96
974	AAAY27587	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 21.	448	100
974	gi12802561	Bovine herpesvirus 4	unknown	74	42
975	AAU16297	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1250.	2420	98
975	AAB94486	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15170.	1761	74
975	AAM94018	Homo sapiens	HELI- Human stomach cancer expressed polypeptide SEQ ID NO 106.	1761	74
976	AAM41264	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6195.	903	99
976	AAM39478	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2623.	903	99
976	AAB43771	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1216.	903	99
977	gi11761611	Homo sapiens.	kinesin-like protein RBKIN1 (RBKIN) mRNA, complete cds, alternatively spliced.	9290	99
977	gi11761613	Homo sapiens	kinesin-like protein RBKIN2 (RBKIN) mRNA, complete cds, alternatively spliced.	9055	98
977	gi12054030	Homo sapiens	mRNA for KINESIN-13A1 (KIN13A gene).	8955	97
978	gi7595802	Mus musculus	ELKL motif kinase 2 short form	188	48
978	gi7595800	Mus musculus	ELKL motif kinase 2 long form	188	48
978	AAM93956	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 4158.	187	48
979	gi16198456	Homo sapiens	Similar to RIKEN cDNA 0610040E02 gene, clone MGC:17973 IMAGE:3919892, mRNA, complete cds.	1050	100
979	gi16740689	Mus musculus	RIKEN cDNA 0610040E02 gene	891	76
979	gi12841315	Mus musculus	putative	891	76
980	gi14714927	Homo sapiens	amino acid transporter system A1, clone MGC:17722 IMAGE:3871101, mRNA, complete cds.	2466	100
980	gi116407	Homo sapiens	amino acid transporter system	2466	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	43		A1 mRNA, complete cds.		
980	AAB93556	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12942.	2459	99
981	AAW75090	Homo sapiens	HUMA- Human secreted protein encoded by gene 34 clone HTEGA81.	507	100
981	AAW75152	Homo sapiens	HUMA- Human secreted protein encoded by gene 34 clone HKMLK44.	507	100
981	AAW75151	Homo sapiens	HUMA- Human secreted protein encoded by gene 34 clone HTEGA81.	507	100
982	AAB94754	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15811.	4658	99
982	gi14042859	Homo sapiens	cDNA FLJ14964 fis, clone PLACE4000581, moderately similar to FIBROPEL LIN 1 PRECURSOR.	4658	99
982	gi11177164	Mus musculus	polydom protein	3880	81
983	AAB65278	Homo sapiens	GETH Human PRO1185 (UNQ599) protein sequence SEQ ID NO:401.	993	100
983	AAM25316	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:831.	993	100
983	AAM23805	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1330.	993	100
984	AAV35996	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 381.	589	66
984	AAB65278	Homo sapiens	GETH Human PRO1185 (UNQ599) protein sequence SEQ ID NO:401.	567	65
984	AAM25316	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:831.	567	65
985	gi16877039	Homo sapiens	hypothetical protein FLJ22688, clone MGC:2438 IMAGE:2819805, mRNA, complete cds.	1952	91
985	gi13325253	Homo sapiens	Similar to hypothetical protein FLJ22688, clone MGC:4098 IMAGE:2819805, mRNA, complete cds.	1952	91
985	gi10439177	Homo sapiens	cDNA: FLJ22688 fis, clone HS111003.	1695	89
986	gi512469	Homo sapiens	H.sapiens HLA-DMA gene.	1285	92
986	gi2181876	Homo sapiens	Human DNA sequence from clone XX-O27 on chromosome 6. Contains the BRD2 gene encoding bromodomain- containing 2 protein, the HLA- DMA gene encoding major histocompatibility complex class II DM alpha, two CpG islands, ESTs, STSs and GSSs, complete sequence.	1285	92
986	gi150303	Homo sapiens	clone MGC:13532	1285	92

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	36		IMAGE:4245221, mRNA, complete cds.		
987	AAB42529	Homo sapiens	CURA- Human ORFX ORF2293 polypeptide sequence SEQ ID NO:4586	2683	99
987	gi12697893	Homo sapiens	mRNA for KIAA1674 protein, partial cds.	2683	99
987	AAM90536	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:18129.	1512	98
988	gi2224539	Homo sapiens	Human mRNA for KIAA0299 gene, partial cds.	9903	99
988	AAAY16588	Homo sapiens	RHON A protein that interacts with presenilins.	4733	97
988	gi7301710	Drosophila melanogaster	CG11754 gene product	3074	43
989	AAW74887	Homo sapiens	HUMA- Human secreted protein encoded by gene 160 clone HCELB21.	203	100
989	AAM24401	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1926.	183	85
989	gi9294003	Arabidopsis thaliana	cytochrome P450-like protein	70	39
990	gi11990420	Homo sapiens	mRNA for MOP-5, complete cds.	4359	95
990	AAB93229	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12223.	3950	99
990	gi7023114	Homo sapiens	cDNA FLJ10833 fis, clone NT2RP4001206, moderately similar to Drosophila melanogaster strawberry notch mRNA.	3950	99
991	gi4337105	Homo sapiens	MSH55 gene, partial cds; and CLIC1, DDAH, G6b, G6c, G5b, G6d, G6e, G6f, BAT5, G5b, CSK2B, BAT4, G4, Apo M, BAT3, BAT2, AIF-1, IC7, LST-1, LTB, TNF, and LTA genes, complete cds.	668	100
991	gi29969	Homo sapiens	Human gene for casein kinase II subunit beta (EC 2.7.1.37).	668	100
991	gi29967	Homo sapiens	Human mRNA for phosphoinositide kinase type II beta subunit (EC 2.7.1.37).	668	100
992	AAAY10840	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	349	100
993	AAM25927	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1442.	824	100
993	AAAY73325	Homo sapiens	INCY- HTRM clone 001106 protein sequence.	820	99
993	AAG03870	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7951.	819	99
994	AAB94043	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14201.	5698	99
994	gi10433976	Homo sapiens	cDNA FLJ12471 fis, clone NT2RM1000894, highly similar to DNA-DIRECTED RNA	5698	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			POLYMERASE I 135 KD POLYPEPTIDE (EC 2.7.7.6).		
994	gi1621594	Mus musculus	second largest subunit of RNA polymerase I	5095	84
995	AAU15880	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 833.	1091	100
995	ABD03345	Homo sapiens	HUMA- Human musculoskeletal system related polypeptide SEQ ID NO 1292.	1091	100
995	gi13879442	Mus musculus	Similar to RIKEN cDNA 2310035M22 gene	1056	93
996	gi6692607	Mus musculus	MGA protein	3446	77
996	gi3043560	Homo sapiens	mRNA for KIAA0518 protein, partial cds.	3272	100
996	AAB94560	Homo sapiens	HELL- Human protein sequence SEQ ID NO:15332.	1933	99
997	gi5771392	Homo sapiens	RAB-like protein 2B (RABL2B) mRNA, complete cds.	718	99
997	gi15928838	Homo sapiens	RAB, member of RAS oncogene family-like 2B, clone MGC:10160 IMAGE:3906749, mRNA, complete cds.	718	99
997	gi7263961	Homo sapiens	Human DNA sequence from clone RP11-395L14. Contains (part of) up to six novel genes or pseudogenes, the gene for a novel forkhead protein similar to FOXD4 (forkhead box D4, FREAC5), the gene for a novel phosphoglucomutase like protein, a pseudogene similar to part of DEAD/H (Asp-Glu-Ala-Asp/His) box (S.cerevisiae CHL1-like helicase), an RPL23A (60S ribosomal protein L23A) pseudogene, the RABL2A gene for RAB-like 2A, the gene for a novel protein similar to small nuclear ribonucleoprotein polypeptide A' (SNRPA1) and the 3' part of the gene for a novel protein similar to acrosin (ACR). Contains ESTs, STSS, GSSs and nine putative CpG islands, complete sequence.	714	97
998	gi10440202	Homo sapiens	cDNA: FLJ23495 fis, clone LNG02228.	2398	99
998	AAU17289	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 854.	487	97
998	AAM92681	Homo sapiens	HUMA- Human digestive system antigen SEQ ID NO: 2030.	487	97
999	gi12653249	Homo sapiens	Similar to CAAX box 1, clone MGC:8471 IMAGE:2821721, mRNA, complete cds.	450	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
999	AAY32207	Homo sapiens	INCY- Human receptor molecule (REC) encoded by Incyte clone 2936050.	429	95
999	gi7576232	Homo sapiens	Human DNA sequence from clone RP4-809E13 on chromosome Xq26.1-27.1. Contains the gene for a putative prenylated protein, two putative prenylated protein pseudogenes, ESTs, STSs, GSSs and three putative CpG islands, complete sequence.	397	87
1000	gi15778556	Homo sapiens	alpha-1-B glycoprotein precursor (A1BG) mRNA, complete cds.	1487	98
1000	gi11877348	Rattus norvegicus	putative alpha 1B-glycoprotein	518	40
1000	AAY64670	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:831.	430	76
1001	AAY87315	Homo sapiens	INCY- Human signal peptide containing protein HSPP-92 SEQ ID NO:92.	2817	100
1001	AAM93793	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3821.	2527	99
1001	gi12804527	Homo sapiens	hypothetical protein FLJ22405, clone MGC:2543 IMAGE:2961594, mRNA, complete cds.	2194	100
1002	gi4160304	Mus musculus	HS1 binding protein 3	1449	75
1002	AAB95801	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18781.	1082	100
1002	gi10436660	Homo sapiens	cDNA FLJ14249 fis, clone OVARC1001200, weakly similar to Mus musculus mRNA for HSI binding protein 3.	1082	100
1003	AAY87315	Homo sapiens	INCY- Human signal peptide containing protein HSPP-92 SEQ ID NO:92.	1837	100
1003	gi12804527	Homo sapiens	hypothetical protein FLJ22405, clone MGC:2543 IMAGE:2961594, mRNA, complete cds.	1837	100
1003	gi10438780	Homo sapiens	cDNA: FLJ22405 fis, clone HRC08294.	1837	100
1004	AAM93793	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3821.	3401	99
1004	gi17390694	Mus musculus	Similar to hypothetical protein FLJ22405	2543	90
1004	AAY87315	Homo sapiens	INCY- Human signal peptide containing protein HSPP-92 SEQ ID NO:92.	2535	100
1005	AAY87327	Homo sapiens	INCY- Human signal peptide containing protein HSPP-104 SEQ ID NO:104.	584	100
1005	AAY59705	Homo sapiens	GEST Secreted protein 51-41-1-F10-FL1.	554	95

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1005	AAAY12865	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO:455.	208	100
1006	AAAY36237	Homo sapiens	HUMA- Human secreted protein encoded by gene 14.	177	100
1007	AAAY87310	Homo sapiens	INCY- Human signal peptide containing protein HSPP-87 SEQ ID NO:87.	370	100
1007	AAG77344	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:8110.	201	79
1007	gi14198125	Homo sapiens	clone MGC:18053 IMAGE:4148889, mRNA, complete cds.	68	61
1008	AAB94108	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14340.	1844	96
1008	AAU04557	Homo sapiens	GETH Human Stra6 homologue, PRO10282.	1844	96
1008	gi13560966	Homo sapiens	STRA6 isoform 1 mRNA, complete cds, alternatively spliced.	1844	96
1009	AAAY38394	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 9.	213	100
1010	gi10432382	Homo sapiens	Human DNA sequence from clone RP4-717123 on chromosome 1p21.2-22.3. Contains ESTs, STSs and GSSs. Contains part of a novel gene for a protein similar to Xenopus laevis Sojo protein, a novel gene and a 60S ribosomal protein L39 (RPL39) pseudogene, complete sequence.	3267	100
1010	gi5690435	Xenopus laevis	nuclear protein Sojo	1386	44
1010	AAG75036	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5800.	557	98
1011	AAG00517	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4598.	160	48
1011	AAO02474	Homo sapiens	IYSE- Human polypeptide SEQ ID NO 16366.	153	45
1011	gi854065	Human herpesvirus 6	U88	145	50
1012	AAAY19561	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	514	100
1012	AAB38157	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 39 SEQ ID NO:96.	70	30
1012	AAU04958	Homo sapiens	GETH Human Interleukin 17 receptor, IL-17RII4.	69	60
1013	AAR15222	Homo sapiens	TEXA Chronic myelogenous leukemia-derived myeloid-related protein.	635	100
1013	gi32402	Homo sapiens	Human mRNA for HP-1, a member of the corticostatin/defensin family.	493	100
1013	gi292363	Homo sapiens	Human neutrophil peptide-1	493	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			gene, complete cds.		
1014	gi6330176	Homo sapiens	mRNA for KIAA1167 protein, partial cds.	4079	99
1014	gi8980841	Rattus norvegicus	GRIP-associated protein 1 long form	3814	92
1014	gi17389263	Mus musculus	Similar to GRIP-associated protein 1	3646	89
1015	gi10443047	Homo sapiens	Human DNA sequence from clone RP11-465L10 on chromosome 20. Contains 10 CpG islands, ESTs, STSs and GSSs. Contains the gene for a novel protein similar to Drosophila CG11399, the gene for a novel C2H2 type zinc finger protein similar to chicken FZF-1, a Ferritin light polypeptide (FTL) pseudogene, the MMP9 gene for matrix metalloproteinase 9 (gelatinase B, 92kD gelatinase, 92kD type IV collagenase) (CLG4B), a novel gene, the SLC12A5 gene for solute carrier family 12, (potassium-chloride transporter) member 5 (KIAA1176) and the 3' end of gene KIAA1637, complete sequence.	6471	99
1015	gi10438918	Homo sapiens	cDNA: FLJ22504 fls, clone HRC11430.	4392	98
1015	gi984814	Gallus gallus	zinc finger protein	2127	58
1016	AAE06077	Homo sapiens	HUMA- Human gene 37 encoded secreted protein HDPCF91, SEQ ID NO:139.	267	100
1016	AAAY87100	Homo sapiens	HUMA- Human secreted protein sequence SEQ ID NO:139.	267	100
1016	gi12718812	Yarrowia lipolytica	ND3 protein	69	48
1017	AAAY86463	Homo sapiens	HUMA- Human gene 47- encoded protein fragment, SEQ ID NO:378.	361	100
1017	AAAY86320	Homo sapiens	HUMA- Human secreted protein HPRBC80, SEQ ID NO:235.	361	100
1017	gi7549633	Arabidopsis thaliana	hypothetical protein	70	31
1018	AAM25384	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:899.	1126	100
1018	AAAY27581	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 15.	774	100
1018	gi13785618	Mus musculus	sideroflexin 4	660	60
1019	gi452890	Cricetulus migratorius	serum amyloid P; SAP; female protein; FP	158	71
1019	gi387051	Cricetulus longicaudatus	FP	157	71
1019	gi347257	Mesocricetus auratus	serum amyloid P component	157	71

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1020	gi17428783	Ralstonia solanacearum	PROBABLE NADP-DEPENDENT OXIDOREDUCTASE OXIDOREDUCTASE PROTEIN	68	29
1020	gi15159226	Agrobacterium tumefaciens str. C58 (Cereon)	AGR_L_1604p	67	28
1020	gi17742500	Agrobacterium tumefaciens str. C58 (Dupont)	succinoglycan biosynthesis protein	67	28
1021	gi16553933	Homo sapiens	cDNA FLJ25217 fis, clone REC08938, highly similar to Oryctolagus cuniculus Na ⁺ /glucose cotransporter-related protein mRNA.	1477	100
1021	AAE06614	Homo sapiens	SAGA Human protein having hydrophobic domain, HP03974.	1394	100
1021	gi15209808	Homo sapiens	unnamed protein product	1394	100
1022	AAV16781	Homo sapiens	GEMY Human secreted protein (clone bh157.7).	1258	100
1022	gi12654011	Homo sapiens	similar to rat nuclear ubiquitous casein kinase 2, clone MGC:5494 IMAGE:3452665, mRNA, complete cds.	1258	100
1022	gi12053624	Homo sapiens	mRNA for NUCKS protein.	1258	100
1023	AAB83246	Homo sapiens	MILL- Human FATP1 SEQ ID NO: 47.	3372	100
1023	AAB83239	Homo sapiens	MILL- Human FATP1 SEQ ID NO: 38.	3372	100
1023	AAB83234	Homo sapiens	MILL- Human FATP1 SEQ ID NO: 32.	3372	100
1024	gi15929904	Homo sapiens	Similar to dolichyl-phosphate mannosyltransferase polypeptide 2, regulatory subunit, clone MGC:21559 IMAGE:4181887, mRNA, complete cds.	366	100
1024	AAV57905	Homo sapiens	INCY- Human transmembrane protein HTMPN-29.	153	100
1024	gi3790363	Homo sapiens	mRNA for DPM2, complete cds.	153	100
1025	AAV25732	Homo sapiens	HUMA- Human secreted protein encoded from gene 22.	212	100
1026	AAG77016	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:7780.	291	100
1026	AAS03193_aal	Homo sapiens	GEHO Human lymphocyte cell surface antigen CD53 cDNA sequence.	116	95
1026	AAV81220_anl	Homo sapiens	GEHO Human CD53 antigen cDNA.	116	95
1027	AAV11848	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID No: 448.	193	100
1027	AAV35956	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO.	193	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			205.		
1027	AAY36098	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 483.	193	100
1028	AAM93942	Homo sapiens	HELL- Human polypeptide, SEQ ID NO: 4126.	1354	99
1028	ABB11422	Homo sapiens	HYSE- Human Zn finger protein homologue, SEQ ID NO:1792.	953	92
1028	gi6467206	Homo sapiens	GIOT-4 mRNA for gonadotropin inducible transcription repressor-4, complete cds.	883	56
1029	gi12835222	Mus musculus	putative	1562	82
1029	gi2636654	Homo sapiens	Human myosin binding protein H (MyBP-H) gene, complete cds.	1206	66
1029	gi1546813	Mus musculus	myosin binding protein H	1203	66
1030	AAY36332	Homo sapiens	HUMA- Human secreted protein encoded by gene 109.	268	100
1030	gi10038917	Buchnera sp. APS	hypothetical protein	76	42
1030	gi7500518	Unknown	hypothetical protein F35E2.7 - Caenorhabditis elegans >	63	38
1031	gi9719409	Homo sapiens	candidate tumor suppressor protein mRNA, complete cds.	2030	99
1031	gi12856514	Mus musculus	putative	922	86
1031	AAU22041	Homo sapiens	HUMA- Human cardiovascular system antigen polypeptide SEQ ID No 815.	703	92
1032	AAR15222	Homo sapiens	TEXA Chronic myelogenous leukaemia-derived myeloid-related protein.	635	100
1032	gi32402	Homo sapiens	Human mRNA for HP-1, a member of the corticostatin/defensin family.	493	100
1032	gi292363	Homo sapiens	Human neutrophil peptide-1 gene, complete cds.	493	100
1033	gi16552502	Homo sapiens	cDNA FLJ32395 fis, clone SKMUS2000117, moderately similar to Homo sapiens MAGEF1 mRNA.	1599	100
1033	gi12659142	Mus musculus	mage-g1	1178	76
1033	gi12857118	Mus musculus	putative	1178	76
1034	AAB49650	Homo sapiens	CURA- Human SEC2 protein sequence SEQ ID 4.	2615	100
1034	gi12226532	Homo sapiens	unnamed protein product	2615	100
1034	gi14714886	Mus musculus	Unknown (protein for IMAGE:3498778)	2343	89
1035	AAM23721	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1246.	2889	100
1035	AAM792	Homo sapiens	HYSE- Human protein SEQ ID	2676	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	29		NO 1891.		
1035	AAB88370	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0106.	1611	100
1036	AAY59657	Homo sapiens	GEST Secreted protein 108-003-5-0-A8-FL.	689	100
1036	gi14495699	Homo sapiens	clone MGC:15961 IMAGE:3538818, mRNA, complete cds.	689	100
1036	gi14424522	Homo sapiens	clone MGC:14327 IMAGE:4298098, mRNA, complete cds.	689	100
1037	AAY27626	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 60.	352	100
1038	AAD18356_aa1	Homo sapiens	INCY- Human lipid metabolism enzyme-5 (LME-5) cDNA.	1748	100
1038	AAB73560	Homo sapiens	MILL- Human lipase 18892.	1748	100
1038	AAE10996	Homo sapiens	INCY- Human lipid metabolism enzyme-5 (LME-5) protein.	1748	100
1039	AAG03475	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7556.	448	100
1039	AAY12861	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO:451.	448	100
1039	gi56760	Rattus norvegicus	neuronal nonacetylcholine binding subunit	75	23
1040	AAY53049	Homo sapiens	GEMY Human secreted protein clone cj378 3 protein sequence SEQ ID NO:104.	463	100
1040	gi13603674	Stellilabium pogonostali	maturase	78	38
1040	gi13603676	Telipogon parvulus	maturase	74	36
1041	AAY41354	Homo sapiens	HUMA- Human secreted protein encoded by gene 47 clone HUF030.	288	100
1041	gi15230414	Arabidopsis thaliana	putative protein	63	43
1042	AAW74777	Homo sapiens	HUMA- Human secreted protein encoded by gene 48 clone HFCA174.	245	100
1042	gi15488920	Homo sapiens	Similar to RIKEN cDNA 2010107G23 gene, clone MGC:9596 IMAGE:3896656, mRNA, complete cds.	245	100
1042	gi12842465	Mus musculus	putative	241	97
1043	gi15278028	Homo sapiens	beta-galactose-3-O-sulfotransferase, 4, clone MGC:15045 IMAGE:3636329, mRNA, complete cds.	2568	100
1043	gi14794524	Homo sapiens	Galbeta1-3GalNAc 3'-sulfotransferase mRNA, complete cds.	2564	99
1043	AAB93892	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13832.	2556	99
1044	gi153215	Homo sapiens	empty spiracles-like protein	1341	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	90		(EMX2) mRNA, complete cds.		
1044	gi13276773	Homo sapiens	mRNA; cDNA DKFZp761M1614 (from clone DKFZp761M1614).	1341	100
1044	gi16549686	Homo sapiens	cDNA FLJ30479 fis, clone BRAWH1000168, highly similar to Homeotic protein emx2.	1336	99
1045	gi16877066	Homo sapiens	clone MGC:24447 IMAGE:4077762, mRNA, complete cds.	362	100
1045	gi16877059	Homo sapiens	clone MGC:24437 IMAGE:4075637, mRNA, complete cds.	362	100
1045	AAY94959	Homo sapiens	GEMY Human secreted protein clone mc300_1 protein sequence SEQ ID NO:124.	204	97
1046	gi2394451	Caenorhabditis elegans	Hypothetical protein ZC178.2	406	30
1046	AAB87575	Homo sapiens	GETH Human PRO1342.	384	38
1046	AAY99408	Homo sapiens	GETH Human PRO1342 (UNQ697) amino acid sequence SEQ ID NO:243.	384	38
1047	gi12053147	Homo sapiens	mRNA; cDNA DKFZp434F1726 (from clone DKFZp434F1726).	1484	98
1047	ABB11739	Homo sapiens	HYSE- Human IF-gamma receptor homologue, SEQ ID NO:2109.	1044	100
1047	AAR04932	Homo sapiens	YEDA Interferon-gamma receptor segment from clone 39 responsible for binding the target.	829	98
1048	AAG73989	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4753.	957	100
1048	AAB58998	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 706.	957	100
1048	AAM89100	Homo sapiens	HUMA- Human immune/hematopoietic antigen SEQ ID NO:16693.	362	67
1049	gi10047333	Homo sapiens	mRNA for KIAA1628 protein, partial cds.	5204	100
1049	gi11862939	Mus musculus	DDM36	4378	89
1049	gi11862941	Mus musculus	DDM36B	4366	88
1050	gi190647	Homo sapiens	Human pregnancy-specific beta-1 glycoprotein (PSG) mRNA, complete cds.	611	72
1050	gi984306	Homo sapiens	Human pregnancy-specific glycoprotein 13 (PSG13) mRNA, complete cds.	606	71
1050	gi190568	Homo sapiens	Human pregnancy-specific beta-1-glycoprotein 11 (PSG11) mRNA, complete cds.	585	64
1051	AAM436	Homo sapiens	HUMA- Human polypeptide	588	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	55		SEQ ID NO 333.		
1051	AAM43588	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 266.	588	100
1051	AAW60043	Homo sapiens	HUTC- Human MHC class I chain-related gene A (MICA) polypeptide.	588	100
1052	gi4877836	Rattus norvegicus	TRP2	524	74
1052	gi11095641	Mus musculus	transient receptor potential channel 2-beta	521	73
1052	gi11095639	Mus musculus	transient receptor potential channel 2-alpha	521	73
1053	AAB94188	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14511.	1209	97
1053	gi10434224	Homo sapiens	cDNA FLJ12623 fis, clone NT2RM4001746.	1209	97
1053	gi12652797	Homo sapiens	clone MGC:5179 IMAGE:2900118, mRNA, complete cds.	1066	88
1054	AAY38389	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 4.	152	90
1054	gi13624635	Euglena viridis	maturase-like protein	63	42
1055	AAY27582	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 16.	320	100
1055	gi13421631	Caulobacter crescentus	conserved hypothetical protein	65	27
1055	gi16124713	Caulobacter crescentus] > [Caulobacter crescentus	conserved hypothetical protein	65	27
1056	AAO8759	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 22651.	613	100
1056	gi10437569	Homo sapiens	cDNA: FLJ21463 fis, clone COL04765.	269	57
1056	AAY45382	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 28.	266	58
1057	AAE05175	Homo sapiens	INCY- Human drug metabolising enzyme (DME-6) protein.	1830	99
1057	AAU12225	Homo sapiens	GETH Human PRO4404 polypeptide sequence.	1830	99
1057	AAU18363	Homo sapiens	HUMA- Human endocrine polypeptide SEQ ID No 318.	1092	95
1058	AAG81274	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:66.	815	96
1058	gi14035856	Homo sapiens	unnamed protein product	815	96
1058	AAG81273	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:64.	652	99
1059	AAY35980	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 229.	581	97
1059	ABB11960	Homo sapiens	HYSE- Human neuroendocrine-specific protein-like homologue, SEQ ID NO:2330.	246	100
1059	AAZ3831	Homo sapiens	PROT- Human transmembrane	240	97

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	9_aa1		protein cDNA clone HP02061.		
1060	gi7959325	Homo sapiens	mRNA for KIAA1529 protein, partial cds.	8481	100
1060	gi12836354	Mus musculus	putative	511	63
1060	AAW03626	Homo sapiens	UYNV Human thyrotropin GPR N-terminal sequence.	236	31
1061	AAV76200	Homo sapiens	HUMA- Human secreted protein encoded by gene 77.	262	100
1061	gi15924350	Staphylococcus aureus subsp. aureus Mu50	oxacillin resistance-related FmtC protein	64	31
1061	gi12483631	Staphylococcus aureus	FmtC	64	31
1062	AAV36270	Homo sapiens	HUMA- Human secreted protein encoded by gene 47.	359	100
1062	gi4996079	Human herpesvirus 6	64% identical to U95 gene of strain U1102 of HHV-6-MCMV JE2 homolog, US22 gene family	68	37
1062	gi5733576	Human herpesvirus 6B	U95	66	37
1063	gi5262748	Rattus norvegicus	Proline rich synapsc associated protein 2	3930	93
1063	gi7381056	Rattus norvegicus	Shank postsynaptic density protein 3a	3895	92
1063	gi13359173	Homo sapiens	mRNA for KIAA1650 protein, partial cds.	3085	100
1064	gi14336749	Homo sapiens	16p13.3 sequence section 6 of 8.	974	99
1064	gi10440021	Homo sapiens	cDNA: FLJ23360 fls, clone HEP15172.	974	99
1064	AAB94188	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14511.	914	78
1065	AAB94188	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14511.	1227	100
1065	gi10434224	Homo sapiens	cDNA FLJ12623 fls, clone NT2RM4001746.	1227	100
1065	gi12652797	Homo sapiens	clone MGC:5179 IMAGE:2900118, mRNA, complete cds.	1084	90
1066	AAV82488	Homo sapiens	NISC- Human L-type amino acid transporter 1 protein sequence SEQ ID NO:2.	2438	94
1066	gi5926732	Homo sapiens	mRNA for L-type amino acid transporter 1, complete cds.	2438	94
1066	gi4426640	Homo sapiens	L-type amino acid transporter subunit LAT1 mRNA, complete cds.	2438	94
1067	AAG81326	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:170.	1135	100
1067	gi14035960	Homo sapiens	unnamed protein product	1135	100
1067	AAV78805	Homo sapiens	PROT- Hydrophobic domain containing protein clone HP10508 protein sequence.	1053	99
1068	gi12052983	Homo sapiens	mRNA; cDNA DKFZp43411610 (from clone DKFZp43411610);	2502	63

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			complete cds.		
1068	AAM79760	Homo sapiens	HYSE- Human protein SEQ ID NO:3406.	2270	61
1068	AAB94388	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14947.	2050	57
1069	gi12655091	Homo sapiens	AD-003 protein, clone MGC:783 IMAGE:3050940, mRNA, complete cds.	546	53
1069	gi6523799	Homo sapiens	adrenal gland protein AD-003 mRNA, complete cds.	536	53
1069	AAM51824	Homo sapiens	BLOW- Human transcription regulator 13.	370	61
1070	gi14328009	Homo sapiens	clone IMAGE:3942111, mRNA, partial cds.	2392	100
1070	gi14585869	Homo sapiens	hypothetical protein SB146	2389	99
1070	gi16041767	Homo sapiens	Similar to NADPH oxidase-related, C2 domain-containing protein, clone MGC:23187 IMAGE:4851468, mRNA, complete cds.	2384	99
1071	ABB12245	Homo sapiens	HYSE- Human CKSR-2 homologue, SEQ ID NO:2615.	256	50
1071	AAZ88239_aa1	Homo sapiens	INCY- Human cytokine signal regulator CKSR-2 encoding cDNA SEQ ID NO:4.	245	51
1071	AAB67667	Homo sapiens	INCY- Amino acid sequence of human cytokine signal regulator 2.	245	51
1072	gi15418997	Homo sapiens	capillary morphogenesis protein-1 mRNA, complete cds.	3015	100
1072	AAB95505	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18062.	2163	99
1072	gi10435718	Homo sapiens	cDNA FLJ13645 fis, clone PLACE1011310, weakly similar to MYOSIN HEAVY CHAIN, GIZZARD SMOOTH MUSCLE.	2163	99
1073	gi15985082	Homo sapiens	unnamed protein product	3158	99
1073	AAU04958	Homo sapiens	GETH Human Interleukin 17 receptor, IL-17RH4.	3148	97
1073	gi6453552	Homo sapiens	mRNA; cDNA DKFZp434N1928 (from clone DKFZp434N1928).	3007	100
1074	AAB93827	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13641.	884	99
1074	AAB23606	Homo sapiens	ALPH- Human secreted protein SEQ ID NO: 12.	884	99
1074	gi10433126	Homo sapiens	cDNA FLJ11790 fis, clone HEMBA1006091.	884	99
1075	gi15777931	Homo sapiens	DGCRK3 gene for G-protein beta subunit like protein, complete cds.	1731	99
1075	gi15082309	Homo sapiens	clone MGC:19898 IMAGE:4548339, mRNA, complete cds.	1731	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1075	gi13359165	Homo sapiens	mRNA for KIAA1645 protein, partial cds.	1731	99
1076	gi16605905	Homo sapiens	unnamed protein product	100	32
1076	gi11611188	Homo sapiens	Human DNA sequence from clone RP4-688G8 on chromosome 20q11.2-12. Contains the gene for a novel protein similar to ribosomal protein S2 (RPS2), a gene encoding a protein similar to basic protease inhibitor chelonianin, a novel gene, the 3' end of a novel gene, ESTs, STSs, GSSs and a CpG island, complete sequence.	94	32
1076	AAY35935	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 184.	83	29
1077	gi16554229	Homo sapiens	cDNA FLJ25436 fls, clone TST08261.	1164	100
1077	gi15082426	Homo sapiens	Similar to RIKEN cDNA 2810055F11 gene, clone MGC:20203 IMAGE:4684687, mRNA, complete cds.	1156	99
1077	gi12858155	Mus musculus	putative	1054	89
1078	gi15559290	Homo sapiens	clone MGC:20275 IMAGE:3842589, mRNA, complete cds.	1917	100
1078	gi15625564	Homo sapiens	WD40- and FYVE-domain containing protein 2 (WDF2) mRNA, complete cds.	1893	99
1078	gi16554204	Homo sapiens	cDNA FLJ25420 fls, clone TST03665.	1380	100
1079	AAB92775	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11256.	3212	99
1079	gi14042169	Homo sapiens	cDNA FLJ14564 fls, clone NT2RM4000229, weakly similar to Gallus gallus actin filament-associated protein (AFAP-110) mRNA.	3212	99
1079	gi15620887	Homo sapiens	mRNA for KIAA1914 protein, partial cds.	1702	100
1080	AAG89172	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 292.	591	100
1080	AAY12532	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO: 197 from WO 9906553.	512	91
1080	AAB87173	Homo sapiens	MILL- Human secreted protein TANGO 402 S22T variant, SEQ ID NO:215.	119	44
1081	gi9963804	Homo sapiens	zinc finger protein ZNF286 (ZNF286) mRNA, complete cds.	574	94
1081	gi14017965	Homo sapiens	mRNA for KIAA1874 protein, partial cds.	517	91
1081	AAU16238	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1191.	362	100

SEQ ID	Ht ID	Species	Description	S score	Percent identity
1082	AAG03810	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7891.	841	99
1082	gi186800	Homo sapiens	Human ribosomal protein L12 mRNA, complete cds.	841	99
1082	gi14198333	Homo sapiens	ribosomal protein L12, clone MGC:9760 IMAGE:3855674, mRNA, complete cds.	841	99
1083	AAY12902	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO:492.	134	100
1084	gi186774	Homo sapiens	Human Kruppel related zinc finger protein (HTF10) mRNA, complete cds.	3082	57
1084	gi2739353	Homo sapiens	DNA from chromosome 19, BAC 33152, complete sequence.	2985	55
1084	AAM79739	Homo sapiens	HYSE- Human protein SEQ ID NO 3385.	2612	53
1085	AAV48125_aa1	Homo sapiens	HUMA- Nucleotide sequence encoding clone HMWGS46 of Prohibitin receptor family.	1354	93
1085	ABB11913	Homo sapiens	HYSE- Human B-cell receptor associated protein homologue, SEQ ID NO:2283.	1354	93
1085	AAY94443	Homo sapiens	UNII Human repressor of estrogen repressor activity (REA) protein.	1354	93
1086	AAG72370	Homo sapiens	YEDA Human OR-like polypeptide query sequence, SEQ ID NO: 2051.	333	100
1086	AAG71453	Homo sapiens	YEDA Human olfactory receptor polypeptide, SEQ ID NO: 1134.	333	100
1086	AAE04556	Homo sapiens	INCY- Human G-protein coupled receptor-12 (GCRC-12) protein.	315	100
1087	AAG81323	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:164.	792	100
1087	gi14035954	Homo sapiens	unnamed protein product	792	100
1087	gi12838799	Mus musculus	putative	564	76
1088	AAE09691	Homo sapiens	HUMA- Human gene 2 encoding novel protein HCOKA10, SEQ ID NO:38.	96	55
1088	AAG76125	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6889.	96	55
1088	AAU16944	Homo sapiens	HUMA- Human novel secreted protein, SEQ ID 185.	96	55
1089	AAY13037	Homo sapiens	GEST Human secreted protein encoded by 5' EST SEQ ID NO: 51.	187	100
1089	AAY36395	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 5.	79	39
1089	gi13096804	Mus musculus	Unknown (protein for IMAGE:3586067)	77	40
1090	gi151478	Mus musculus	Spred-2	2098	92

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	77				
1090	gi15147875	Mus musculus	Spred-1	1101	52
1090	AAU17415	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 980.	1029	98
1091	gi16551762	Homo sapiens	cDNA FLJ31812 fis, clone NT2R12009406, moderately similar to Homo sapiens rec mRNA.	804	97
1091	gi6682873	Homo sapiens	rec mRNA, complete cds.	453	55
1091	gi7230612	Rattus norvegicus	small rec	451	56
1092	gi12224885	Homo sapiens	mRNA; cDNA DKFZp76111011 (from clone DKFZp76111011).	241	94
1092	gi3880560	Caenorhabditis elegans	Similarity to Yeast E1-E2 ATPase (SW:YED1_YEAST), contains similarity to Pfam domain: PF00122 (E1-E2 ATPase), Score=102.4, E-value=2.7e-28, N=4-cDNA EST yk5f9.5 comes from this gene-cDNA EST yk10d12.5 comes from this gene-cDNA EST yk5f9.3 comes from this gene-cDNA EST yk10d12.3 comes from this gene-cDNA EST yk40h11.5 comes from this gene-cDNA EST yk131g11.3 comes from this gene-cDNA EST yk131g11.5 comes from this gene-cDNA EST yk133d6.5 comes from this gene-cDNA EST yk318f2.3 comes from this gene-cDNA EST yk318f2.5 comes from this gene-cDNA EST yk122a12.5 comes from this gene-cDNA EST yk248b3.5 comes from this gene-cDNA EST yk260b8.5 comes from this gene-cDNA EST yk286h5.5 comes from this gene-cDNA EST yk356g1.5 comes from this gene	135	50
1092	gi16516658	Homo sapiens	ORF for hypothetical protein.	129	52
1093	AAB40996	Homo sapiens	CURA- Human ORFX ORF760 polypeptide sequence SEQ ID NO:1520.	349	40
1093	gi16516658	Homo sapiens	ORF for hypothetical protein.	349	40
1093	gi10436963	Homo sapiens	cDNA: FLJ20986 fis, clone CAE01156.	349	40
1094	gi16904208	Homo sapiens	very large G protein-coupled receptor 1b (VLGR1) mRNA, complete cds.	1014	96
1094	gi169042	Mus musculus	very large G protein-coupled	808	78

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	10		receptor 1		
1094	gi12044471	Homo sapiens	mRNA; cDNA DKFZp761P0710 (from clone DKFZp761P0710); complete cds.	173	27
1095	gi12483902	Rattus norvegicus	zinc finger protein HIT-10	1545	47
1095	AAB95862	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18929.	1234	50
1095	gi10436789	Homo sapiens	cDNA FLJ14345 fls, clone THYRO1001189, weakly similar to ZINC FINGER PROTEIN 91.	1234	50
1096	AAB50963	Homo sapiens	GETH Human PRO1286 protein.	466	100
1096	AAU12421	Homo sapiens	GETH Human PRO1286 polypeptide sequence.	466	100
1096	AAU09179	Homo sapiens	GETH Human PRO1268 polypeptide.	466	100
1097	AAE12023	Homo sapiens	INCY- Human G-protein coupled receptor, GCRC-2.	2849	98
1097	AAG68126	Homo sapiens	FARB Human 7TM-GPCR protein sequence SEQ ID NO:6.	2824	98
1097	gi17512539	Mus musculus	Unknown (protein for MGC:29266)	2183	73
1098	AAB94108	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14340.	273	100
1098	AAU04558	Homo sapiens	GETH Human Stra6 homologue, PRO19578.	273	100
1098	AAU04557	Homo sapiens	GETH Human Stra6 homologue, PRO10282.	273	100
1099	AAU12382	Homo sapiens	GETH Human PRO792 polypeptide sequence.	137	32
1099	AAB24416	Homo sapiens	GETH Human PRO792 protein sequence SEQ ID NO:155.	137	32
1099	AAB24055	Homo sapiens	GETH Human PRO792 protein sequence SEQ ID NO:31.	137	32
1100	gi6330422	Homo sapiens	mRNA for KIAA1202 protein, partial cds.	4913	99
1100	gi12314062	Homo sapiens	Human DNA sequence from clone RP11-119E20 on chromosome Xp11.21-11.23. Contains part of the gene for KIAA1202 protein, ESTs, STSs and GSSs, complete sequence.	4696	99
1100	gi15421201	Homo sapiens	SHAP-A (SHAP) mRNA, partial cds, alternatively spliced.	3845	99
1101	gi10834607	Homo sapiens	cadherin 20 (CDH20) mRNA, complete cds.	4170	99
1101	gi4101751	Mus musculus	cadherin 7 precursor	4032	96
1101	gi854635	Xenopus laevis	F-cadherin	3251	78
1102	AAV36310	Homo sapiens	HUMA- Human secreted protein encoded by gene 87.	250	100
1103	gi16551423	Homo sapiens	cDNA FLJ31547 fls, clone NT2R12001010, weakly similar to FATTY ACYL-COA	853	98

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			HYDROLASE PRECURSOR, MEDIUM CHAIN (EC 3.1.2.14).		
1103	gi2641986	Mesocricetus auratus	carboxylesterase precursor	438	50
1103	AAD00680_aa1	Homo sapiens	INCV- Human Hydrolase protein-5 (HYDRL-5) encoding cDNA.	428	52
1104	AAE04836	Homo sapiens	SUGE- Human SGP018 phosphatase polypeptide.	4915	95
1104	gi12718835	Homo sapiens	unknown mRNA.	3350	99
1104	AAB40919	Homo sapiens	CURA- Human ORFX ORF683 polypeptide sequence SEQ ID NO:1366.	718	79
1105	gi13492650	Rattus norvegicus	potassium channel beta subunit KChIP4	1284	99
1105	gi11527318	Mus musculus	calseinin-like protein	1281	99
1105	gi14091336	Homo sapiens	KCHIP4.1 (KCNIP4) mRNA, complete cds, alternatively spliced.	1278	99
1106	AAAY13126	Homo sapiens	GBST Human secreted protein encoded by 5' EST SEQ ID NO: 140.	160	96
1107	gi4126593	Cyprinus carpio	complement C3-S	156	29
1107	gi4126589	Cyprinus carpio	complement C3-H1	148	26
1107	gi305335	Cavia porcellus	complement C3 protein (GPC3) precursor	146	50
1108	gi11244873	Homo sapiens	PR-domain-containing protein 16 (PRDM16) mRNA, complete cds.	6646	99
1108	gi12697895	Homo sapiens	mRNA for KIAA1675 protein, partial cds.	3570	99
1108	gi545408	human, leukemic cell line SKH1, mRNA Mutant, 5938 nt. [Homo sapiens	AML1-EVI-1=AML1-EVI-1 fusion protein {rearranged translocation}	3181	53
1109	gi10732815	Homo sapiens	concentrative Na+-nucleoside cotransporter hCNT3 (CNT3) mRNA, complete cds.	3609	100
1109	gi10732817	Mus musculus	concentrative Na+-nucleoside cotransporter mCNT3	2872	78
1109	gi15489379	Mus musculus	solute carrier family 28 (sodium-coupled nucleoside transporter), member 3	2859	77
1110	gi17865150	Plasmodium berghei	cysteine repeat modular protein 3 PbCRM3	97	30
1110	gi6650711	Giardia intestinalis	variant-specific surface protein VSP1267-2	96	29
1110	gi861294	Caenorhabditis elegans	F35D2.4 gene product	94	31
1111	gi7959177	Homo sapiens	mRNA for KIAA1458 protein, partial cds.	2993	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1111	AAB94791	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15907.	1037	100
1111	AAU01101	Homo sapiens	HUMA- Gene 38 Human secreted protein homologous amino acid sequence.	842	99
1112	gi17390760	Mus musculus	RIKEN cDNA 2610205H19 gene	664	99
1112	gi12848292	Mus musculus	putative	664	99
1112	gi203072	Rattus sp.	0-44 protein	661	98
1113	gi14250319	Homo sapiens	clone IMAGE:3448367, mRNA, partial cds.	2143	93
1113	gi14582773	Homo sapiens	sumo/sentrin-specific protease	2138	93
1113	gi17026032	Macaca fascicularis	hypothetical protein	2068	89
1114	AAB93777	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13502.	1064	99
1114	AAM41387	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6318.	1064	99
1114	AAM39601	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2746.	1064	99
1115	gi15277240	Homo sapiens	genomic DNA, chromosome 6p21.3, HLA Class I region, section 17/20.	2256	100
1115	gi11875405	Homo sapiens	HZFw1 protein mRNA, complete cds.	2251	99
1115	gi11875407	Homo sapiens	HZFw2 protein mRNA, complete cds.	1733	99
1116	AAB95726	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18602.	1644	99
1116	AAB95109	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17089.	1644	99
1116	gi14042078	Homo sapiens	cDNA FLJ14510 fis, clone NT2RM1000623, weakly similar to RIBONUCLEASE INHIBITOR.	1644	99
1117	gi14009346	Homo sapiens	nGAP-like protein (AF9q34) mRNA, complete cds.	5475	98
1117	gi15277525	Rattus norvegicus	DOC2/DAB2 interactive protein	5006	96
1117	gi12698051	Homo sapiens	mRNA for KIAA1743 protein, partial cds.	3024	98
1118	AAB65211	Homo sapiens	GETH Human PRO1152 (UNQ582) protein sequence SEQ ID NO:216.	1937	99
1118	AAB68883	Homo sapiens	INCY- Human RECAP polypeptide, SEQ ID NO: 13.	1937	99
1118	AAU28183	Homo sapiens	HYSE- Novel human secretory protein, Seq ID No 352.	1937	99
1119	AAG02197	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6278.	522	99
1119	AAU17224	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 789.	485	100
1119	AAU17597	Homo sapiens	HUMA- Novel signal transduction pathway protein,	449	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			Seq ID 1162.		
1120	gi179636	Homo sapiens	cytoplasmic phosphotyrosyl protein phosphatase (clone type 1) complete cds.	717	88
1120	gi1147812	Homo sapiens	Human red cell-type low molecular weight acid phosphatase (ACP1) gene, exon 6 and 7, complete cds.	717	88
1120	gi5759131	Rattus norvegicus	low molecular weight protein tyrosine phosphatase isoform A	647	76
1121	gi7670466	Mus musculus	unnamed protein product	255	61
1121	gi3413810	Mus musculus	Bassoon	105	25
1121	gi50715	Mus musculus	myosin heavy chain	103	26
1122	gi10440335	Homo sapiens	cDNA: FLJ23594 fis, clone L.NG14867.	1513	100
1122	gi14602889	Homo sapiens	clone MGC:13119 IMAGE:4100726, mRNA, complete cds.	702	53
1122	gi14249827	Homo sapiens	clone MGC:10992 IMAGE:3637387, mRNA, complete cds.	702	53
1123	ABB11200	Homo sapiens	HYSE- Human Kupffer cell receptor homologue, SEQ ID NO:1570.	1838	99
1123	gi15489066	Mus musculus	Kupffer cell c-type lectin receptor	1000	45
1123	gi1669360	Mus musculus	Kupffer cell receptor	1000	45
1124	AAAY30847	Homo sapiens	HUMA- Human secreted protein encoded from gene 37.	239	100
1124	gi14596027	Arabidopsis thaliana	Unknown protein	70	45
1124	gi9758957	Arabidopsis thaliana	contains similarity to unknown protein-gb AAAF64546.1-gene_id:MRB17.15	70	45
1125	AAAY01390	Homo sapiens	HUMA- Secreted protein encoded by gene 8 clone HTXD188.	301	100
1125	gi15639942	Treponema pallidum	dicarboxylate transporter (dctM)	63	38
1126	AAG68189	Homo sapiens	GENO- Cytosolic thyroid hormone-binding protein SEQ ID NO:105.	152	78
1126	gi35505	Homo sapiens	H.sapiens M gene for M1-type and M2-type pyruvate kinase.	152	78
1126	gi338827	Homo sapiens	Human TCB gene encoding cytosolic thyroid hormone-binding protein, complete cds.	152	78
1127	gi4261689	Homo sapiens	complement factor B mRNA, complete cds.	3976	97
1127	gi297569	Homo sapiens	H.sapiens mRNA for complement factor B.	3976	97
1127	AAAX041	Homo sapiens	UNI W Human Factor B	3972	97

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	25 aa1		encoding cDNA.		
1128	AAB50012	Homo sapiens	PHAA Wild-type human alpha7 ligand gated ion channel.	1794	100
1128	AAB82690	Homo sapiens	TEWE- Nicotinic acetylcholine receptor alpha7.	1794	100
1128	AAB24088	Homo sapiens	GETH Human PRO2145 protein sequence SEQ ID NO:77.	1794	100
1129	gi537329	Homo sapiens	Human (clone pHAI2-12) alpha-2 collagen type IV (COL4A2) mRNA, 3' end.	3756	99
1129	gi1022323	Mus musculus	collagen alpha-2(IV) chain	3480	87
1129	gi556299	Mus musculus	alpha-2 type IV collagen	3477	87
1130	gi15011489	Tetrahymena thermophila	heme maturase	68	29
1131	AAB50964	Homo sapiens	GETH Human PRO1313 protein.	926	100
1131	AAB47290	Homo sapiens	GETH PRO1313 polypeptide.	926	100
1131	AAB24431	Homo sapiens	GETH Human PRO1313 protein sequence SEQ ID NO:216.	926	100
1132	gi12855307	Mus musculus	putative	2919	89
1132	gi15559525	Homo sapiens	Similar to RIKEN cDNA 4932416D09 gene, clone IMAGE:4578228, mRNA, partial cds.	2523	99
1132	AAY54052	Homo sapiens	PHAA An angiogenesis-associated protein which binds plasminogen.	1435	62
1133	AAY13084	Homo sapiens	GEST Human secreted protein encoded by 5' EST SEQ ID NO: 98.	127	62
1133	gi10440468	Homo sapiens	mRNA for FLJ00070 protein, partial cds.	75	41
1133	gi4558640	Homo sapiens	chromosome 19, cosmid R27516, complete sequence.	74	44
1134	AAM80275	Homo sapiens	HYSE- Human protein SEQ ID NO 3921.	1510	99
1134	AAM79291	Homo sapiens	HYSE- Human protein SEQ ID NO 1953.	1500	99
1134	gi16877449	Homo sapiens	hypothetical protein MGC20781, clone MGC:21670 IMAGE:3885455, mRNA, complete cds.	1367	100
1135	gi7710869	Homo sapiens	Human DNA sequence from clone RP11-31M2 on chromosome 9p23-24.3. Contains (part of) the gene for a novel protein similar to the GLI family of zinc finger proteins, STSs, GSSs and two putative CpG islands, complete sequence.	1629	100
1135	AAM25473	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:988.	861	98
1135	gi14486069	Drosophila melanogaster	Zn finger transcription factor lame duck	699	63

SEQ ID	Hlt ID	Species	Description	S score	Percent Identity
1136	AAR05562	Homo sapiens	DANA- Laminin -binding protein encoded by insert from J9 lambda gt10 phage.	1327	94
1136	gi307105	Homo sapiens	Human colon carcinoma laminin-binding protein mRNA, complete cds.	1327	94
1136	gi16307602	Homo sapiens	laminin receptor 1 (67kD, ribosomal protein SA), clone MGC:17122 IMAGE:3446816, mRNA, complete cds.	1327	94
1137	AAB43884	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1329.	734	98
1137	gi179281	Homo sapiens	ATP synthase beta subunit precursor (ATPSB) gene, complete cds.	734	98
1137	gi16741373	Homo sapiens	Similar to ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide, clone MGC:5231 IMAGE:2900336, mRNA, complete cds.	734	98
1138	AAG01468	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5549.	282	98
1138	AAG01467	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5548.	206	100
1138	AAB43884	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1329.	206	100
1139	gi179279	Homo sapiens	Human ATP synthase beta subunit gene, exons 8-10.	757	69
1139	gi12845667	Mus musculus	putative	744	68
1139	gi28940	Homo sapiens	Human mRNA for F1-ATPase beta subunit (F-1 beta).	742	69
1140	AAB43884	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1329.	1124	89
1140	gi179281	Homo sapiens	ATP synthase beta subunit precursor (ATPSB) gene, complete cds.	1124	89
1140	gi16741373	Homo sapiens	Similar to ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide, clone MGC:5231 IMAGE:2900336, mRNA, complete cds.	1124	89
1141	AAW54079	Homo sapiens	TEXA Homo sapiens BARD1 sequence.	4101	100
1141	gi1710175	Homo sapiens	Human BRCA1-associated RING domain protein (BARD1) mRNA, complete cds.	4101	100
1141	AAW54081	Homo sapiens	TEXA Homo sapiens BARD1 P553 sequence.	4097	99
1142	AAW54089	Homo sapiens	TEXA Homo sapiens BARD1 MR658C sequence.	394	100
1142	AAW54088	Homo sapiens	TEXA Homo sapiens BARD1 MS761N sequence.	394	100
1142	AAW540	Homo sapiens	TEXA Homo sapiens BARD1	394	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	87		MQ564H sequence.		
1143	gi17432241	Homo sapiens	MSTP027 (MST027) mRNA, complete cds.	730	100
1143	gi16041688	Homo sapiens	hypothetical protein FLJ21661, clone MGC:16816 IMAGE:3922036, mRNA, complete cds.	730	100
1143	gi14039831	Homo sapiens	elongation factor G2 (EFG2) mRNA, complete cds; nuclear gene for mitochondrial product.	730	100
1144	AAM40729	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5660.	271	98
1144	AAM38943	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2088.	271	98
1144	AAY12325	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO:356.	271	98
1145	gi17390320	Homo sapiens	clone MGC:9678 IMAGE:3846678, mRNA, complete cds.	872	100
1145	gi1200202	Homo sapiens	clone 022f05 My030 protein mRNA, complete cds.	872	100
1145	AAY36068	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 453.	844	97
1146	gi17390320	Homo sapiens	clone MGC:9678 IMAGE:3846678, mRNA, complete cds.	504	95
1146	gi1200202	Homo sapiens	clone 022f05 My030 protein mRNA, complete cds.	504	95
1146	AAY36068	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 453.	495	94
1147	AAB61617	Homo sapiens	PROT- Human protein HP10688.	1488	100
1147	gi12578471	Homo sapiens	unnamed protein product	1488	100
1147	AAY02781	Homo sapiens	HUMA- Human secreted protein.	1146	77
1148	gi1654002	Homo sapiens	Hsapiens mRNA for Sop2p-like protein.	572	99
1148	gi12805063	Mus musculus	actin related protein 2/3 complex, subunit 1A (41 kDa)	567	97
1148	gi12667258	Rattus norvegicus	suppressor of profilin/p41 of actin-related complex 2/3	567	97
1149	AAB95258	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17435.	3560	100
1149	gi10434740	Homo sapiens	cDNA FLJ12957 fls, clone NT2RP2005531, weakly similar to PROTEIN 4.1.	3560	100
1149	gi10047161	Homo sapiens	mRNA for KIAA1548 protein, partial cds.	2487	100
1150	AAR99844	Homo sapiens	SUME Human natural killer cell, cell surface mol. NKG7.	552	71
1150	AAQ86384_aa1	Homo sapiens	ASAN/ G-CSF stimulated human myelocytic cell cDNA.	307	65
1150	AAQ98551_aa1	Homo sapiens	ASAN/ Granulocyte colony stimulating factor (G-CSF)-	307	65

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			inducible myelocyte gene.		
1151	AAM23781	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1306.	620	89
1151	AAM23837	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1362.	481	100
1151	AAY64747	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:908.	471	98
1152	AAB65277	Homo sapiens	GETH Human PRO1187 (UNQ601) protein sequence SEQ ID NO:399.	656	100
1152	AAY66754	Homo sapiens	GETH Membrane-bound protein PRO1187.	656	100
1152	AAB24077	Homo sapiens	GETH Human PRO1187 protein sequence SEQ ID NO:55.	656	100
1153	gi8331757	Homo sapiens	X28 region near ALD locus containing dual specificity phosphatase 9 (DUSP9), ribosomal protein L18a (RPL18a), Ca2+/calmodulin-dependent protein kinase I (CAMK1), creatine transporter (CRTTR), CDM protein (CDM), adrenoleukodystrophy protein (ALD), plexin-related protein (PLXB3), muscle-specific serine kinase (MSSK), NAD-isocitrate dehydrogenase (IDH), translocon-associated protein delta (TRAP), and LU1 protein (LU1) genes, complete cds; and CCp pseudogene, complete sequence.	1747	92
1153	AAW88436	Homo sapiens	INCY- Disease associated protein kinase DAPK-5.	1667	88
1153	gi3135197	Rattus norvegicus	Ca2+/calmodulin-dependent protein kinase 1 beta 2	1661	87
1154	AAG67548	Homo sapiens	LEXI- Amino acid sequence of a human transporter protein.	413	98
1154	gi16588684	Homo sapiens	anion transporter/exchanger-8 (SLC26A8) mRNA, complete cds.	413	98
1154	gi15341554	Homo sapiens	putative anion transporter (SLC26A8) mRNA, complete cds.	413	98
1155	AAB64953	Homo sapiens	ROSE/ Human secreted protein sequence encoded by gene 12 SEQ ID NO:131.	995	98
1155	gi12854324	Mus musculus	putative	781	65
1155	gi15217075	Homo sapiens	glioma pathogenesis-related protein (RTVP1) mRNA, complete cds.	443	40
1156	AAB26256	Homo sapiens	UNLO Wild-type human Rac1 protein.	508	100
1156	gi8574038	Homo sapiens	rac1 gene.	508	100
1156	gi2326206	Homo sapiens	Rac3 (RAC3) mRNA, complete cds.	508	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1157	gi10439853	Homo sapiens	cDNA: FLJ23235 fls, clone CAS04980.	1536	100
1157	AAW85596	Homo sapiens	MERI Human GABA-A receptor theta subunit.	73	35
1157	gi15978247	Yersinia pestis	putative membrane protein	73	27
1158	gi10439853	Homo sapiens	cDNA: FLJ23235 fls, clone CAS04980.	1347	90
1158	AAW85596	Homo sapiens	MERI Human GABA-A receptor theta subunit.	73	35
1158	gi15978247	Yersinia pestis	putative membrane protein	73	27
1159	gi11933149	Homo sapiens	mRNA for 6-phosphofructo-2-kinase heart isoform, complete cds.	2452	100
1159	gi3090419	Homo sapiens	pfkfb2 gene, exons 1 to 15.	2329	99
1159	gi3090421	Homo sapiens	mRNA for 6-phosphofructo-2-kinase.	2319	98
1160	gi6177785	Homo sapiens	mRNA for HKR1, partial cds.	3083	99
1160	gi13325427	Homo sapiens	clone IMAGE:3928207, mRNA, partial cds.	2388	99
1160	gi487783	Homo sapiens	Human zinc finger protein ZNF133.	1592	54
1161	gi14583859	Homo sapiens	hypothetical protein SB138	1558	98
1161	AAB94641	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15526.	1543	100
1161	AAG64403	Homo sapiens	SHAN- Human paneth cell enhanced expression-like protein.	1543	100
1162	AAV53641	Homo sapiens	CHIR A bone marrow secreted protein designated BMS42.	2182	99
1162	gi9663153	Homo sapiens	partial mRNA for transport-secretion protein 2.2, (TTS-2.2 gene).	2179	98
1162	gi9663151	Homo sapiens	partial mRNA for transport-secretion protein 2.1 (TTS-2.1 gene).	2179	98
1163	AAM93360	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2919.	3300	100
1163	gi16878206	Homo sapiens	hypothetical protein DKFZp434J037, clone MGC:29812 IMAGE:5088037, mRNA, complete cds.	3300	100
1163	gi12053281	Homo sapiens	mRNA; cDNA DKFZp434J037 (from clone DKFZp434J037); complete cds.	3300	100
1164	AAG81282	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:82.	3032	100
1164	AAU17102	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 667.	3032	100
1164	gi17529989	Homo sapiens	oxysterol-binding protein-like protein OSBPL9 (OSBPL9) mRNA, complete cds.	3032	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1165	gi14627121	Homo sapiens	Human DNA sequence from clone RP5-824F16 on chromosome 20 Contains the 5' end of the ANGPT4 gene for angiotensin 4, part of the gene for a novel protein similar to mouse thrombospondin type 1 domain protein R-spondin, ESTs, STSs, GSSs and a CpG island, complete sequence.	230	64
1165	gi16605378	Mus musculus	unnamed protein product	226	42
1165	gi12850680	Mus musculus	putative	226	42
1166	AAF84171_aa1	Homo sapiens	CHUG- Human OATP-B coding sequence.	3573	97
1166	AAZ92403_aa1	Homo sapiens	SCHE cDNA encoding human DC-PGT.	3573	97
1166	AAC61883_aa1	Homo sapiens	CHIR cDNA encoding a human secreted protein.	3573	97
1167	gi30224	Homo sapiens	H.sapiens CRP mRNA for C-reactive protein.	327	100
1167	gi30213	Homo sapiens	H.sapiens mRNA for C-reactive protein.	327	100
1167	gi181068	Homo sapiens	Human C-reactive protein gene, complete cds.	327	100
1168	AAH76194_aa1	Homo sapiens	MILL- Human seven-transmembrane protein 31945 coding sequence.	3429	99
1168	AAB85767	Homo sapiens	MILL- Human seven-transmembrane protein 31945 sequence.	3429	99
1168	gi16551933	Homo sapiens	cDNA FLJ31951 fis, clone: NT2RP7007177, weakly similar to Homo sapiens multiple membrane spanning receptor TRC8 mRNA.	3429	99
1169	AAB60299	Homo sapiens	MILL- Human aminopeptidase 17867.	5048	99
1169	AAE04879	Homo sapiens	INCY- Human protease protein-6 (PRTS-6).	5048	99
1169	gi11065900	Homo sapiens	aminopeptidase mRNA, complete cds.	5048	99
1170	gi12844136	Mus musculus	putative	700	52
1170	AAA99905_aa1	Homo sapiens	GETH cDNA encoding human protein PRO846.	662	94
1170	AAB65300	Homo sapiens	GETH Human PRO846 protein sequence SEQ ID NO:517.	662	94
1171	gi12653943	Homo sapiens	clone MGC:2742 IMAGE:2822914, mRNA, complete cds.	3104	100
1171	AAG01236	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5317.	522	94
1171	gi1304387	Saccharomyces cerevisiae var. diastaticus	glucoamylase	196	22

SEQ ID	HIT ID	Species	Description	S score	Percent identity
1172	gi13430408	Homo sapiens	BTBD2 protein mRNA, complete cds.	2705	99
1172	gi17026062	Mus musculus	glucose signal repressing protein	1946	77
1172	gi13430406	Homo sapiens	BTBD1 protein mRNA, complete cds.	1937	76
1173	gi17226121	Homo sapiens	F-box protein (FBG4) mRNA, complete cds.	1503	100
1173	gi16553918	Homo sapiens	cDNA FLJ25205 fis, clone REC05844, highly similar to Mus musculus F-box protein FBX17 mRNA.	1503	100
1173	gi15214527	Homo sapiens	Similar to f-box only protein 17, clone MGC:9379 IMAGE:3864760, mRNA, complete cds.	1503	100
1174	AAB88373	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0109.	2158	100
1174	AAB93215	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12194.	2158	100
1174	AAB93142	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12045.	2158	100
1175	AAB88373	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0109.	3646	95
1175	AAB93215	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12194.	3646	95
1175	gi14042571	Homo sapiens	cDNA FLJ14791 fis, clone NT2RP4001064, weakly similar to SYNAPTONEMAL COMPLEX PROTEIN SC65.	3646	95
1176	AAB36392	Homo sapiens	CHUG- Human tumour suppressor Gros1-S protein SEQ ID NO:4.	3861	99
1176	gi11127638	Homo sapiens	GROS1-L protein mRNA, complete cds.	3861	99
1176	AAB88373	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0109.	3847	99
1177	gi10438539	Homo sapiens	cDNA: FLJ22233 fis, clone HRC02016.	2015	100
1177	AAE04892	Homo sapiens	INCY- Human transporter and ion channel-5 (TRICH-5) protein.	2009	99
1177	gi13925661	Mus musculus	sodium/calcium exchanger protein	1708	84
1178	AAB65192	Homo sapiens	GETH Human PRO839 (UNQ472) protein sequence SEQ ID NO:167.	366	100
1178	AAG81432	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:382.	366	100
1178	AAY66669	Homo sapiens	GETH Membrane-bound protein PRO839.	366	100
1179	gi599683	Bos taurus	Cleavage and Polyadenylation specificity factor (CPSF) 100kD subunit	4034	98

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1179	gi15489017	Mus musculus	cleavage and polyadenylation specific factor 2, 100kD subunit	3993	97
1179	gi2331036	Mus musculus	cleavage and polyadenylation specificity factor	3993	97
1180	AAZ34648_aa1	Homo sapiens	ZYMO Human growth factor zalpha5 cDNA.	2182	91
1180	AAZ45852_aa1	Homo sapiens	COMP- Human liver angiotensin-like growth factor DNA sequence.	2182	91
1180	AAA49716_aa1	Homo sapiens	GETH Human PRO179 cDNA clone DNA16451-1078.	2182	91
1181	AAH23183_aa1	Homo sapiens	ISIS- Human macrophage migration inhibitory factor encoding DNA.	564	94
1181	AAB60325	Homo sapiens	KIRI Human wild-type glycosylation-inhibiting factor (GIF).	564	94
1181	AAB85343	Homo sapiens	ISIS- Human macrophage migration inhibitory factor.	564	94
1182	gi7263938	Homo sapiens	mRNA for sodium-glucose cotransporter (SGLT2 gene).	3408	100
1182	gi5679464	Homo sapiens	Human DNA sequence from clone RP1-90G24 on chromosome 22 Contains the RPPL2 gene for RET finger protein-like 2, a Immunoglobulin Lambda Light Chain C region (IGLC) pseudogene, the gene for SAAT1 (low affinity sodium glucosylcotransporter (sodium:solute symporter family)) and a Cleavage and Polyadenylation Specific Factor CPSF 160 kD subunit pseudogene. Contains ESTs, GSSs and three putative CpG islands, complete sequence.	3408	100
1182	AAY31221	Homo sapiens	KOEP/ Human SAAT1 protein.	3390	99
1183	gi1362301	Homo sapiens	Similar to Zinc transporter 2, clone MGC:11303 IMAOB:3948165, mRNA, complete cds.	1659	100
1183	gi1256378	Rattus norvegicus	zinc transporter ZnT-2	1186	67
1183	gi1763374	Mus musculus	ZnT-3	721	44
1184	AAB92636	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10951.	1480	81
1184	gi7022185	Homo sapiens	cDNA FLJ10250 fis, clone HEMBB1000973, moderately similar to Mus musculus schlafen3 mRNA.	1480	81
1184	AAM40357	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3502.	1479	81
1185	AAB95351	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17641.	3148	99

SEQ ID	Hit ID	Species	Description	S score	Percent Identity
1185	gi10435075	Homo sapiens	cDNA FLJ13170 fis, clone NT2RP3003809, weakly similar to SAV PROTEIN.	3148	99
1185	gi12654321	Homo sapiens	clone MGC:5347 IMAGE:2985725, mRNA, complete cds.	3106	99
1186	gi10567590	Homo sapiens	sodium bicarbonate cotransporter-like protein mRNA, partial cds.	5645	100
1186	gi13447747	Homo sapiens	sodium bicarbonate cotransporter NBC4a (NBC4) mRNA, complete cds.	5486	99
1186	gi15419575	Homo sapiens	sodium bicarbonate cotransporter NBC4c (NBC4) mRNA, complete cds, alternatively spliced.	5382	98
1187	gi10185824	Homo sapiens	SEBOX (SEBOX) gene, complete cds.	1209	100
1187	gi10092160	Mus musculus	SEBOX	581	63
1187	gi10185826	Rattus norvegicus	SEBOX	565	63
1188	AAM43540	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 218.	246	94
1188	AAM40468	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5399.	246	94
1188	AAM38682	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 1827.	246	94
1189	gi10438135	Homo sapiens	cDNA: FLJ21924 fis, clone HEP04086.	3703	99
1189	AAM68038	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28344.	700	100
1189	AAM55656	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 27761.	700	100
1190	gi14388555	Macaca fascicularis	hypothetical protein	1240	96
1190	AAG93258	Homo sapiens	NISC- Human protein HP10582.	979	100
1190	gi12854823	Mus musculus	putative	882	72
1191	AAB94545	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15296.	2344	99
1191	gi10435228	Homo sapiens	cDNA FLJ13273 fis, clone QVARC1001010.	2344	99
1191	gi12860475	Mus musculus	putative	1811	83
1192	gi10440442	Homo sapiens	mRNA for FLJ00057 protein, partial cds.	3033	99
1192	gi12082303	Mus musculus	DNA helicase B	1697	61
1192	gi2635211	Bacillus subtilis	similar to conjugation transfer protein	114	22
1193	AAE10445	Homo sapiens	BIOJ Human transmembrane protein (TMP).	2286	100
1193	AAAY862	Homo sapiens	HUMA- Human secreted protein	1889	85

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	30		HKFBC53, SEQ ID NO:145.		
1193	AAW68002	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 69.	798	96
1194	gi17511729	Homo sapiens	hypothetical protein FLJ12598, clone MGC:31807 IMAGE:5552964, mRNA, complete cds.	1180	100
1194	AAB94162	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14456.	1173	99
1194	gi10434183	Homo sapiens	cDNA FLJ12598 fis, clone NT2RM4001384.	1173	99
1195	gi7243069	Homo sapiens	mRNA for KIAA1344 protein, partial cds.	4137	100
1195	AAB94073	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14262.	1839	99
1195	gi10434025	Homo sapiens	cDNA FLJ12501 fis, clone NT2RM2001681.	1839	99
1196	AAB08894	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 4 SEQ ID NO:51.	235	67
1196	gi335913	Vesicular stomatitis virus	glycoprotein	71	30
1196	gi2960093	Mycobacterium tuberculosis H37Rv	hypothetical protein Rv3669	71	30
1197	AAM94312	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 2970.	1211	98
1197	gi8439396	HERV-H/env62	envelope protein	763	36
1197	gi4959382	Homo sapiens	human endogenous retrovirus HERV-H19 pol protein (pol) gene, partial cds; env protein (env) gene, complete cds; and 3' LTR, complete sequence.	757	36
1198	gi14017381	Homo sapiens	tumor endothelial marker 8 precursor (TEM8) mRNA, complete cds.	1512	100
1198	gi10437939	Homo sapiens	cDNA: FLJ21776 fis, clone HEP00171.	1512	100
1198	gi15987505	Mus musculus	tumor endothelial marker 8 precursor	1484	97
1199	AAB65270	Homo sapiens	GETH Human PRO1158 (UNQ588) protein sequence SEQ ID NO:375.	609	100
1199	AAB87559	Homo sapiens	GETH Human PRO1158.	609	100
1199	AAAY66747	Homo sapiens	GETH Membrane-bound protein PRO1158.	609	100
1200	AAM41380	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6311.	844	87
1200	AAAY01785	Homo sapiens	INCY- Human ubiquitin-conjugating enzyme HUB1-1.	818	87
1200	AAAY25341	Homo sapiens	PROS- Human NCE-2 protein.	818	87
1201	AAW748	Homo sapiens	HUMA- Human secreted protein	197	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	99		encoded by gene 172 clone HODCW06.		
1201	gi13794493	Guillardia theta	hypothetical protein	67	36
1202	AAS07940_aa1	Homo sapiens	AREN- Human cDNA encoding G-protein coupled receptor, hRUP13.	2087	92
1202	AAS12583_aa1	Homo sapiens	FARB DNA encoding human serotonin-like G protein-coupled receptor (5-HT-GPCR).	2087	92
1202	AAD19579_aa1	Homo sapiens	INCY- Human G-protein coupled receptor, GCRC-3 cDNA.	2087	92
1203	AAS07940_aa1	Homo sapiens	AREN- Human cDNA encoding G-protein coupled receptor, hRUP13.	2318	100
1203	AAS12583_aa1	Homo sapiens	FARB DNA encoding human serotonin-like G protein-coupled receptor (5-HT-GPCR).	2318	100
1203	AAD19579_aa1	Homo sapiens	INCY- Human G-protein coupled receptor, GCRC-3 cDNA.	2318	100
1204	AAM93612	Homo sapiens	HELL- Human polypeptide, SEQ ID NO: 3436.	3475	100
1204	gi14043111	Homo sapiens	Similar to ubiquitin associated and SH3 domain containing, A, clone MGC:15437 IMAGE:2958242, mRNA, complete cds.	3412	100
1204	gi16304176	Homo sapiens	nm23-phosphorylated unknown substrate mRNA, complete cds.	2759	100
1205	AAB01424	Homo sapiens	MILL- Human TANGO 213.	1264	100
1205	AAM25735	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1250.	1066	100
1205	AAY76267	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 11.	1066	100
1206	AAW74939	Homo sapiens	HUMA- Human secreted protein encoded by gene 49 clone HAGB117.	211	100
1207	gi15126559	Mus musculus	Similar to Cd63 antigen	504	99
1207	gi14198088	Mus musculus	Cd63 antigen	504	99
1207	gi5410605	Mus musculus	tetraspanin membrane protein CD63	504	99
1208	AAB95148	Homo sapiens	HELL- Human protein sequence SEQ ID NO:17173.	3590	99
1208	gi11990593	Homo sapiens	organic anion transporter polypeptide-related protein 4 (OATPRP4) mRNA, complete cds.	3515	99
1208	AAB49147	Homo sapiens	BRIM Human organic anion transport protein RP4 protein.	3503	99
1209	AAB73381	Homo sapiens	NANF- Human gas vesicle protein homologue hGvpT-b.	1866	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1209	gi12005509	Homo sapiens	HT025 mRNA, complete cds.	1866	100
1209	gi10440236	Homo sapiens	cDNA: FLJ23518 fis, clone LNG04878.	1600	100
1210	gi12053021	Homo sapiens	mRNA; cDNA DKFZp434L0714 (from clone DKFZp434L0714); complete cds.	4230	99
1210	AAG64376	Homo sapiens	BIOD- Human II aminoacyl-tRNA synthetase 75.	3517	99
1210	gi10439991	Homo sapiens	cDNA: FLJ23339 fis, clone HEP13401.	3010	99
1211	AAB58941	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 649.	1188	99
1211	gi12005918	Homo sapiens	CDA016 mRNA, complete cds.	1176	98
1211	AAB36615	Homo sapiens	INCY- Human FLEXHT-37 protein sequence SEQ ID NO:37.	862	98
1212	gi10437194	Homo sapiens	cDNA: FLJ21148 fis, clone CAS09413.	2238	96
1212	gi12845648	Mus musculus	putative	1808	80
1212	gi1226033	Saccharomyces cerevisiae	unknown	259	25
1213	gi10437194	Homo sapiens	cDNA: FLJ21148 fis, clone CAS09413.	2203	91
1213	gi12845648	Mus musculus	putative	1774	76
1213	gi1226033	Saccharomyces cerevisiae	unknown	249	25
1214	gi6630992	Danio rerio	NCC receptor protein 1	418	42
1214	gi12711629	Oreochromis niloticus	nonspecific cytotoxic cell receptor protein	389	41
1214	gi6630998	Ictalurus punctatus	NCC receptor protein 1	361	40
1215	AAC84382_aa1	Homo sapiens	MILL- Human TANGO 209 polypeptide coding sequence.	2430	100
1215	AAS14576_aa1	Homo sapiens	ELIL Human cDNA encoding cysteine-rich secreted protein hCRSP1.	2430	100
1215	AAB48107	Homo sapiens	MILL- Human TANGO 209 polypeptide.	2430	100
1216	ABB11880	Homo sapiens	HYSE- Human lipocortin homologue, SEQ ID NO:2250.	1091	99
1216	AAB43443	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:888.	1071	99
1216	AAR22402	Homo sapiens	BIOS Human lipocortin.	1050	99
1217	ABB11880	Homo sapiens	HYSE- Human lipocortin homologue, SEQ ID NO:2250.	1511	100
1217	AAB43443	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:888.	1511	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1217	AAAY08412	Homo sapiens	UYMC- Human p-40/annexin I protein.	1511	100
1218	ABB11880	Homo sapiens	HYSE- Human lipocortin homologue, SEQ ID NO:2250.	967	100
1218	AAB43443	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:888.	967	100
1218	AAAY08412	Homo sapiens	UYMC- Human p-40/annexin I protein.	967	100
1219	gi4106984	Homo sapiens	Human DNA from chromosome 19-specific cosmid R30923, genomic sequence, complete sequence.	2992	100
1219	AAB42792	Homo sapiens	CURA- Human ORFX ORF2556 polypeptide sequence SEQ ID NO:5112.	2967	99
1219	gi14603176	Homo sapiens	Similar to RIKEN cDNA 2410153K17 gene, clone MGC:19595 IMAGE:3840843, mRNA, complete cds.	2432	100
1220	AAG81443	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:404.	492	100
1220	gi15080220	Homo sapiens	Similar to hypothetical protein, MGC:7764, clone MGC:20548 IMAGE:3607345, mRNA, complete cds.	492	100
1220	gi14056194	Homo sapiens	unnamed protein product	492	100
1221	AAE05183	Homo sapiens	INCY- Human drug metabolising enzyme (DME-14) protein.	2258	100
1221	AAAY91348	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 3 SEQ ID NO:69.	2258	100
1221	gi1185452	Homo sapiens	Human cytochrome P450 monooxygenase CYP2J2 mRNA, complete cds.	932	44
1222	AAE05183	Homo sapiens	INCY- Human drug metabolising enzyme (DME-14) protein.	1344	100
1222	AAAY91348	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 3 SEQ ID NO:69.	1344	100
1222	gi1185452	Homo sapiens	Human cytochrome P450 monooxygenase CYP2J2 mRNA, complete cds.	548	44
1223	AAW34618	Homo sapiens	IMUT- Human C3 protein mutant DV-7N.	597	34
1223	AAW34617	Homo sapiens	IMUT- Human C3 protein mutant DV-6.	597	34
1223	AAW34616	Homo sapiens	IMUT- Human C3 protein mutant CV-5.	597	34
1224	gi17390000	Homo sapiens	Similar to RIKEN cDNA 5730455O13 gene, clone MGC:24718 IMAGE:4278022, mRNA, complete cds.	1693	100
1224	AAB417	Homo sapiens	CURA- Human ORFX	1166	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	53		ORF1517 polypeptide sequence SEQ ID NO:3034.		
1224	gi12857019	Mus musculus	putative	1036	87
1225	AAG00903	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4984.	294	100
1225	gi339697	Homo sapiens	thymosin beta-10 gene, 3'end.	169	97
1225	gi339687	Homo sapiens	Human thymosin beta-10 mRNA, complete cds.	169	97
1226	AAY91386	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 41 SEQ ID NO:107.	558	100
1226	gi9992893	Homo sapiens	phosphoinositol 3-phosphate binding protein-1 (PEPPI) mRNA, complete cds.	75	40
1226	gi10190744	Homo sapiens] > [Homo sapiens	pleckstrin homology domain- containing, family A (phosphoinositide binding specific) member 4; phosphoinositol 3-phosphate binding protein-1	75	40
1227	AAY60008	Homo sapiens	META- Human endometrium tumour EST encoded protein 68.	2286	100
1227	AAW74797	Homo sapiens	HUMA- Human secreted protein encoded by gene 68 clone HKIXR69.	2286	100
1227	gi5762305	Mus musculus	COPI protein	2268	99
1228	AAG89292	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 412.	119	71
1228	AAM93724	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3677.	119	71
1228	gi12803669	Homo sapiens	CDK4-binding protein p34SEI1, clone MGC:3465 IMAGE:3613213, mRNA, complete cds.	119	71
1229	AAY91370	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 25 SEQ ID NO:91.	293	100
1230	gi12004583	Mus musculus	unknown	2566	81
1230	gi12836562	Mus musculus	putative	2541	80
1230	AAB41860	Homo sapiens	CURA- Human ORFX ORF1624 polypeptide sequence SEQ ID NO:3248.	1401	100
1231	AAG76080	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6844.	300	84
1231	AAG01347	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5428.	300	84
1231	AAG01346	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5427.	300	84
1232	ABB11655	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2025.	2233	99
1232	gi15929748	Mus musculus	Unknown (protein for IMAGE:422865)	1826	81

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1232	gi12852713	Mus musculus	putative	1815	81
1233	AAB27977	Homo sapiens	HUMA- Human secreted protein BLAST search protein SEQ ID NO: 131.	290	96
1233	AAAY13458	Homo sapiens	UYRQ Amino acid sequence of human Fe65.	290	96
1233	gi3924936	Homo sapiens	Fe65 protein gene, exons 3 through 14 and partial cds.	290	96
1234	AAO07768	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 21660.	294	100
1234	AAAY02775	Homo sapiens	HUMA- Human secreted protein encoded by gene 12 clone HFTCU19.	288	98
1234	gi7297226	Drosophila melanogaster	CG4497 gene product	67	42
1235	AAB94220	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14581.	3191	99
1235	gi10434288	Homo sapiens	cDNA FLJ12661 fis, clone NT2RM4002189, weakly similar to GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3).	3191	99
1235	gi14018379	Schizosaccharomyces pombe	hypothetical protein; sequence orphan; low similarity to glycoamylases and other cell surface proteins; contains ~250-270 copies of a 13 AA repeat, NSSTPTSSIL	355	26
1236	AAU03593	Homo sapiens	INCY- Human DNA modification protein, DNAMP-8.	4977	98
1236	gi6063137	Mus musculus	F-box protein FBX18	4406	92
1236	AAB94200	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14538.	4039	99
1237	AAB01395	Homo sapiens	INCY- Neuron-associated protein.	1507	95
1237	AAU20529	Homo sapiens	HUMA- Human secreted protein, Seq ID No 521.	1156	86
1237	gi12843076	Mus musculus	putative	644	95
1238	gi16550822	Homo sapiens	cDNA FLJ31400 fis, clone NT2NE1000185, weakly similar to UDP-N-ACETYLGLUCOSAMINE--PEPTIDE N-ACETYLGLUCOSAMINYLTRANSFERASE 110 KDA SUBUNIT (EC 2.4.1.-).	1999	96
1238	gi13604167	Homo sapiens	ARG99 mRNA, complete cds.	781	100
1238	gi6721161	Arabidopsis thaliana	putative O-linked GlcNAc transferase	372	27
1239	AAAY73414	Homo sapiens	GEMY Human secreted protein clone yb101_1 protein sequence SEQ ID NO:50.	472	100
1239	gi719065	Chlamydia	conserved hypothetical protein	70	28

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	6	muridarum			
1239	gi3328759	Chlamydia trachomatis	hypothetical protein	68	28
1240	AAW88615	Homo sapiens	HUMA- Secreted protein encoded by gene 82 clone HNGBT31.	525	97
1240	AAV84040	Homo sapiens	USGO Amino acid sequence of cancer associated polypeptide CHI-9a11-2.	71	30
1240	gi14860975	human herpesvirus 2	DNA polymerase	70	36
1241	gi11121483	Homo sapiens	mRNA for calyntenin-2 (CS2 gene).	5080	100
1241	gi11121506	Mus musculus	calyntenin-2	4733	94
1241	gi11558248	Gallus gallus	calyntenin-1 protein	2962	57
1242	AAM93376	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2951.	1034	100
1242	AAW78151	Homo sapiens	HUMA- Human secreted protein encoded by gene 26 clone HT3BE24.	1034	100
1242	AAV29865	Homo sapiens	GEMY Human secreted protein clone pe213 1.	1034	100
1243	gi10439594	Homo sapiens	cDNA: FLJ23033 fis, clone LNG02005.	2982	99
1243	AAB41396	Homo sapiens	CURA- Human ORFX ORF1160 polypeptide sequence SEQ ID NO:2320.	2486	100
1243	gi12853018	Mus musculus	putative	2002	88
1244	AAV12252	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO: 565.	321	92
1244	AAU16332	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1285.	320	92
1244	AAU15873	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 826.	320	92
1245	AAE04878	Homo sapiens	INCY- Human protease protein-5 (PRTS-5).	1542	100
1245	gi13543681	Homo sapiens	clone MGC:14793 IMAGE:4047601, mRNA, complete cds.	1524	99
1245	AAB47527	Homo sapiens	MILL- Ubiquitin hydrolase-like protein - long form.	1499	100
1246	gi14043523	Homo sapiens	clone IMAGE:4098694, mRNA, partial cds.	1991	97
1246	gi12060822	Homo sapiens	serologically defined breast cancer antigen NY-BR-16 mRNA, complete cds.	1991	97
1246	gi12963869	Mus musculus	gene trap ankyrin repeat containing protein	1980	96
1247	AAB54357	Homo sapiens	HUMA- Human pancreatic cancer antigen protein sequence SEQ ID NO:809.	301	100
1247	AAV48600	Homo sapiens	META- Human breast tumour-associated protein 61.	285	98
1247	gi156129	Bacillus	BH0396-unknown conserved	63	32

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	59	halodurans] > [Bacillus halodurans	protein		
1248	AAM405 66	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5497.	379	46
1248	AAM387 80	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 1925.	379	46
1248	gi653960 6	Homo sapiens	metastasis suppressor protein mRNA, complete cds.	379	46
1249	gi139925 24	Homo sapiens	mRNA for type II alpha phosphatidylinositol 4-kinase gene.	2546	100
1249	gi131119 89	Homo sapiens	Similar to hypothetical protein FLJ111105, clone MGC:4395 IMAGE:2905670, mRNA, complete cds.	2546	100
1249	gi136607 55	Rattus norvegicus	55 kDa type II phosphatidylinositol 4-kinase	2409	94
1250	AAB954 25	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17833.	4871	100
1250	gi104354 87	Homo sapiens	cDNA FLJ13465 fis, clone PLACE1003493, weakly similar to ENDOTHELIAL CELL MULTIMERIN PRECURSOR.	4871	100
1250	AAAY363 00	Homo sapiens	HUMA- Human secreted protein encoded by gene 77.	2472	98
1251	gi644932 6	Mus musculus	retinoic acid-responsive protein HAIR-62	83	68
1251	gi647869 4	Soybean mosaic virus	P1 protease	67	33
1252	gi124078 29	Homo sapiens	netrin 4 precursor (NTN4) mRNA, complete cds.	3361	99
1252	AAG664 49	Homo sapiens	GEHO Human beta-netrin.	3347	99
1252	gi111200 48	Homo sapiens	beta-netrin mRNA, complete cds.	3347	99
1253	gi168780 83	Homo sapiens	enolase 3, (beta, muscle), clone MGC:29581 IMAGE:4902149, mRNA, complete cds.	558	94
1253	gi34789	Homo sapiens	H.sapiens mRNA for muscle specific enolase (MSE) (EC 4.2.1.11).	555	94
1253	gi31170	Homo sapiens	Human ENO3 mRNA for beta-enolase (EC 4.2.1.11).	551	93
1254	AAAY078 95	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 44.	537	100
1254	gi171320 82	Nostoc sp. PCC 7120	ORF_ID:alr2988-hypothetical protein	69	38
1255	AAB937 52	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13419.	1000	99
1255	gi104327 85	Homo sapiens	cDNA FLJ11515 fis, clone HEMBA1002241, weakly similar to PROLIFERATING-CELL NUCLEOLAR ANTIGEN P120.	1000	99
1255	gi156801 85	Homo sapiens	Similar to RIKEN cDNA 2810405F18 gene, clone	875	99

SEQ ID	HIT ID	Species	Description	S score	Percent identity
			MGC:22960 IMAGE:4865283, mRNA, complete cds.		
1256	gi14160880	Homo sapiens	PKCI-1-related HIT protein mRNA, complete cds.	827	100
1256	gi13650128	Homo sapiens	HIT-17kDa mRNA, complete cds.	827	100
1256	AAM25739	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1254.	806	94
1257	gi12849277	Mus musculus	putative	793	93
1257	gi6425295	Caenorhabditis elegans	predicted using Genefinder--contains similarity to Pfam domain: PF00023 (Ank repeat), Score=71.3, E-value=6.5e-18, N=2	200	40
1257	gi4335756	Arabidopsis thaliana	putative ankyrin	195	44
1258	AAB87609	Homo sapiens	GETH Human PRO1890.	1307	99
1258	AAB73309	Homo sapiens	UROG- Human C-type lectin transmembrane antigen FC-LECTIN, SEQ ID NO:2.	1307	99
1258	AAU12441	Homo sapiens	GETH Human PRO1890 polypeptide sequence.	1307	99
1259	AAY05368	Homo sapiens	UYPR- Human HCMV inducible gene protein, SEQ ID NO 4.	1682	97
1259	AAY07036	Homo sapiens	LUDW- Breast cancer associated antigen precursor sequence.	1682	97
1259	gi9956035	Homo sapiens	clone CDABP0047 mRNA sequence.	1682	97
1260	AAB18968	Homo sapiens	INCY- Amino acid sequence of a human transmembrane protein.	1132	100
1260	gi15530313	Homo sapiens	clone MGC:2853 IMAGE:2987806, mRNA, complete cds.	1132	100
1260	gi13937595	Homo sapiens	Similar to RIKEN cDNA 1810017F10 gene, clone MGC:2583 IMAGE:3140820, mRNA, complete cds.	1132	100
1261	AAB18968	Homo sapiens	INCY- Amino acid sequence of a human transmembrane protein.	926	85
1261	gi15530313	Homo sapiens	clone MGC:2853 IMAGE:2987806, mRNA, complete cds.	926	85
1261	gi13937595	Homo sapiens	Similar to RIKEN cDNA 1810017F10 gene, clone MGC:2583 IMAGE:3140820, mRNA, complete cds.	926	85
1262	AAB94434	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15054.	1629	100
1262	gi17512103	Homo sapiens	hypothetical protein FLJ13044, clone MGC:20950 IMAGE:4577143, mRNA, complete cds.	1629	100
1262	gi104348	Homo sapiens	cDNA FLJ13044 fis, clone	1629	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	74		NT2RP3001355, weakly similar to TRICARBOXYLATE TRANSPORT PROTEIN PRECURSOR.		
1263	gi178640	Homo sapiens	Human angiotensinogen mRNA, complete cds.	2366	96
1263	gi1197497	Homo sapiens	H.sapiens angiotensinogen gene exon 2 (and joined cds).	2366	96
1263	AAB67350	Homo sapiens	UTAH Human angiotensinogen protein.	2363	96
1264	gi2401261	Homo sapiens	HLA-C gene (HLA-Cw*0701 allele), complete cds.	1099	98
1264	gi15277217	Homo sapiens	genomic DNA, chromosome 6p21.3, HLA Class I region, section 7/20.	1099	98
1264	gi1478197	Homo sapiens	H.sapiens mRNA for human leukocyte antigen C alpha chain.	1099	98
1265	gi386775	Homo sapiens	Human MHC class I HLA-B*8 chain gene (A1,2; B5,8), complete cds.	1033	92
1265	gi2401259	Homo sapiens	HLA-B gene (HLA-B*0801 allele), complete cds.	1033	92
1265	gi15277216	Homo sapiens	genomic DNA, chromosome 6p21.3, HLA Class I region, section 6/20.	1033	92
1266	AAM23760	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1285.	937	100
1266	gi458664	Homo sapiens	Human MHC class I antigen HLA-B (HLA-B-0704 allele) mRNA, complete cds.	937	100
1266	gi307221	Homo sapiens	Human MHC HLA protein (allele B7) complete cds.	937	100
1267	gi32181	Homo sapiens	H.sapiens HLA-Bw57 gene.	977	88
1267	gi307222	Homo sapiens	Human MHC HLA protein, allele B57, complete cds.	977	88
1267	gi6746371	Homo sapiens	HLA class I histocompatibility antigen B-57 (HLA-B57) mRNA, complete cds.	971	88
1268	gi14970574	Homo sapiens	HLA-A gene for MHC class I antigen, allele HLA-A*68011, exons 1-8.	1801	94
1268	gi1724034	Homo sapiens	Human HLA class I A locus antigen A*68new mRNA, complete cds.	1796	93
1268	gi6138770	Homo sapiens	HLA-A gene for MHC Class I antigen, A*68 allele, exons 1-8.	1792	93
1269	gi307225	Homo sapiens	Human MHC HLA protein, allele A25, complete cds.	1160	96
1269	gi14250359	Homo sapiens	clone MGC:17191 IMAGE:4157200, mRNA, complete cds.	1160	96
1269	gi1526976	Homo sapiens	H.sapiens mRNA for human leukocyte antigen, HLA-A25.	1152	96
1270	gi6453365	Homo sapiens	mRNA for human leukocyte antigen B (HLA-B gene, B*1501102N allele).	314	88
1270	AAV647	Homo sapiens	GEST Human 5' EST related	148	81

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	49		polypeptide SEQ ID NO:910.		
1270	AAP70155	Homo sapiens	BEHW Sequence encoded by genomic DNA encoding human histocompatibility antigen HLA-B 27.	138	62
1271	gi825674	Homo sapiens	H.sapiens gene encoding HLA-Cw6, exons 1-3.	1120	99
1271	gi297097	Homo sapiens	H.sapiens mRNA for HLA-Cw*0602.	1120	99
1271	gi1944480	Homo sapiens	mRNA for HLA-Cw*0602, partial cds.	1120	99
1272	gi2225890	Homo sapiens	Human HLA-A26null allele, complete cds.	977	85
1272	gi487909	Homo sapiens	mRNA for HLA-A11 antigen A11.1, complete cds.	847	94
1272	gi4468256	Homo sapiens	mRNA for MHC class I antigen, allele A*1103.	847	94
1273	gi860968	Homo sapiens	Human HLA-A1 gene.	1122	100
1273	gi8250245	Homo sapiens	HLA-A*0101 gene for MHC class I antigen, exons 1-8.	1122	100
1273	gi386893	Homo sapiens	Human MHC class I HLA-A1 chain gene (A1,2; B8,5), complete cds.	1122	100
1274	AAB94486	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:15170.	2982	100
1274	AAM94018	Homo sapiens	HEL1- Human stomach cancer expressed polypeptide SEQ ID NO 106.	2982	100
1274	gi14042496	Homo sapiens	cDNA FLJ14750 fis, clone NT2RP3002948, weakly similar to RING CANAL PROTEIN.	2982	100
1275	gi12053277	Homo sapiens	mRNA; cDNA DKFZp434B227 (from clone DKFZp434B227); complete cds.	2242	99
1275	gi10440305	Homo sapiens	cDNA: FLJ23571 fis, clone LNG12303.	2124	94
1275	gi11611603	Macaca fascicularis	hypothetical protein	2064	90
1276	gi8918932	Mus musculus	unnamed protein product	2826	95
1276	gi6329812	Homo sapiens	mRNA for KIAA1130 protein, partial cds.	2716	100
1276	AAS14595_aa1	Homo sapiens	MILL- Human cDNA encoding a novel glycosyltransferase 33877.	1606	58
1277	AAB90676	Homo sapiens	GEMY Human BV141_2 protein sequence SEQ ID 28.	400	98
1277	AAW58985	Homo sapiens	GEMY Homo sapiens adult brain clone BV141_2 encoded protein.	201	100
1277	gi2950481	Schizosaccharomyces pombe	hypothetical protein	71	34
1278	AAV14455	Homo sapiens	HUMA- Human secreted protein encoded by gene 45 clone HCFBJ91.	284	100
1279	AAB85885	Homo sapiens	HEL1- Human adenylate kinase 3 (AK3)-like protein.	135	78

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1279	AAB93487	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12786.	135	78
1279	AAB93066	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11883.	135	78
1280	AAB65242	Homo sapiens	GETH Human PRO1291 (UNQ659) protein sequence SEQ ID NO:291.	1378	100
1280	AAB87555	Homo sapiens	GETH Human PRO1291.	1378	100
1280	AAY66719	Homo sapiens	GETH Membrane-bound protein PRO1291.	1378	100
1281	AAB95682	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18481.	1675	96
1281	gi14041989	Homo sapiens	cDNA FLJ14456 fis, clone HEMBB1001915, moderately similar to UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 64E (EC 3.1.2.15).	1675	96
1281	gi7295436	Drosophila melanogaster	Ubp64E gene product	892	71
1282	AAW67841	Homo sapiens	HUMA- Human secreted protein encoded by gene 35 clone HOABG65.	500	100
1282	AAY12238	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO: 551.	423	100
1282	AAY11953	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID No: 553.	276	94
1283	AAM25958	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1473.	484	78
1283	AAW67869	Homo sapiens	HUMA- Human secreted protein encoded by gene 63 clone HHGDB72.	484	78
1283	AAY02285	Homo sapiens	GEMY Secreted protein clone k232_2x polypeptide sequence.	467	78
1284	gi29963	Homo sapiens	Human gene for creatine kinase B (EC 2.7.3.2).	162	71
1284	gi180570	Homo sapiens	Human creatine kinase isozyme CK-B gene, exon 8.	162	71
1284	gi180555	Homo sapiens	Human creatine kinase-B mRNA, complete cds.	162	71
1285	gi12849820	Mus musculus	putative	1170	71
1285	AAM25389	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:904.	790	98
1285	AAY24917	Homo sapiens	INCY- Human phosphatase HPA-2.	550	39
1286	gi12849820	Mus musculus	putative	1456	85
1286	AAY24917	Homo sapiens	INCY- Human phosphatase HPA-2.	798	48
1286	gi8979825	Homo sapiens	Human DNA sequence from clone RP4-776F14 on chromosome 20p12.2-13. Contains the 5' end of the FKBP1A gene for FK506- binding protein 1A (12kD), the	798	48

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			gene for P47 protein, part of a novel member of the PTPNS (protein tyrosine phosphatase, non-receptor type substrate 1) gene family, ESTs, STSs, GSSs and two CpG islands, complete sequence.		
1287	gi3171908	Homo sapiens	mRNA for DnaJ protein.	659	100
1287	gi16041837	Homo sapiens	DnaJ (Hsp40) homolog, subfamily A, member 2, clone MGC:9488 IMAGE:3922477, mRNA, complete cds.	659	100
1287	gi15278395	Homo sapiens	Similar to DnaJ (Hsp40) homolog, subfamily A, member 2, clone MGC:819 IMAGE:3505399, mRNA, complete cds.	659	100
1288	gi15741221	Homo sapiens	gene overexpressed in astrocytoma mRNA, complete cds.	3391	99
1288	gi13544035	Homo sapiens	clone IMAGE:3535476, mRNA, partial cds.	2095	100
1288	gi16878187	Homo sapiens	Similar to gene overexpressed in astrocytoma, clone MGC:29809 IMAGE:5017710, mRNA, complete cds.	2079	100
1289	AAY92719	Homo sapiens	GENZ Human polycystin.	20114	99
1289	gi904223	Homo sapiens	polycystic kidney disease 1 protein (PKD1) mRNA, complete cds.	20114	99
1289	AAW23830	Homo sapiens	DEKR- Human PKD1 protein.	20111	99
1290	AAY55965	Homo sapiens	SUGE- Full length human ZC4 protein.	1906	100
1290	AAY55934	Homo sapiens	SUGE- Human ZC4 protein.	1808	100
1290	gi2780173	Homo sapiens	Human DNA sequence from PAC 82/11 and cosmid U134E6 on chromosome Xq22. Contains NIK like and Thyroxin-binding globulin precursor (T4-binding globulin, TBG) genes, ESTs and STSs.	1588	95
1291	gi10438063	Homo sapiens	cDNA: FLJ21868 fis, clone HEP02432.	1605	99
1291	gi15277443	Mus musculus	Unknown (protein for MGC:19083)	1379	84
1291	AAB42953	Homo sapiens	CURA- Human ORFX ORF2717 polypeptide sequence SEQ ID NO:5434.	522	100
1292	gi7959263	Homo sapiens	mRNA for KIAA1501 protein, partial cds.	1824	100
1292	ABB17488	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 6145.	984	100
1292	AAB979	Homo sapiens	SHAN- Human G-protein	965	57

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	11		activating protein 129 SEQ ID NO:2.		
1293	gi10443048	Homo sapiens	Human DNA sequence from clone RP11-465L10 on chromosome 20. Contains 10 CpG islands, ESTs, STSs and GSSs. Contains the gene for a novel protein similar to Drosophila CG11399, the gene for a novel C2H2 type zinc finger protein similar to chicken FZF-1, a Ferritin light polypeptide (FTL) pseudogene, the MMP9 gene for matrix metalloproteinase 9 (gelatinase B, 92kD gelatinase, 92kD type IV collagenase) (CLG4B), a novel gene, the SLC12A5 gene for solute carrier family 12, (potassium-chloride transporter) member 5 (KIAA1176) and the 3' end of gene KIAA1637, complete sequence.	3567	90
1293	AAH28226_aa1	Homo sapiens	PF1Z Nucleotide sequence of matrix metalloproteinase-9.	3556	90
1293	AAB20491	Homo sapiens	SMIK Human matrix metalloproteinase-9 (MMP-9).	3556	90
1294	AAH28226_aa1	Homo sapiens	PF1Z Nucleotide sequence of matrix metalloproteinase-9.	2375	100
1294	AAB20491	Homo sapiens	SMIK Human matrix metalloproteinase-9 (MMP-9).	2375	100
1294	AAB84611	Homo sapiens	PF1Z Amino acid sequence of matrix metalloproteinase-9.	2375	100
1295	AAG04088	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8169.	601	91
1295	gi897761	Homo sapiens	H.sapiens mRNA for protein phosphatase 5.	450	92
1295	gi4558638	Homo sapiens	chromosome 19, BAC 82621 (CIT-B-139a18), complete sequence.	450	92
1296	AAAY64786	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:947.	200	100
1296	gi16264094	Sinorhizobium meliloti	HYPOTHETICAL PROTEIN	63	35
1297	gi12698013	Homo sapiens	mRNA for KIAA1734 protein, partial cds.	3889	100
1297	gi10438694	Homo sapiens	cDNA: FLJ22346 fis, clone HRC06158.	3877	99
1297	AAB94354	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14875.	2421	99
1298	gi14575679	Homo sapiens	hemiscentin mRNA, complete cds.	10314	89
1298	gi16551993	Homo sapiens	cDNA FLJ31995 fis, clone NT2RF7009236, weakly similar to BASEMENT MEMBRANE-SPECIFIC HEPARAN	4274	91

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			SULFATE PROTEOGLYCAN CORE PROTEIN PRECURSOR		
1298	gi13872813	Homo sapiens	partial mRNA for fibulin-6 (FIBL-6 gene)	2907	99
1299	gi548084	Rattus norvegicus	olfactory cyclic nucleotide-gated channel	2811	93
1299	gi538129	Rattus norvegicus	cyclic nucleotide gated cation channel	2811	93
1299	gi908824	Bos taurus	alpha subunit of CNG-channel expressed in bovine testis and retinal cone	1576	53
1300	AAB41963	Homo sapiens	CURA- Human ORFX ORF1727 polypeptide sequence SEQ ID NO:3454	514	100
1300	gi15529703	Homo sapiens	importin 9 mRNA, complete cds.	514	100
1300	gi15186758	Mus musculus	RANBP9 isoform 2	514	100
1301	gi10505349	Homo sapiens	regulator of G-protein signaling (RGS8) mRNA, complete cds.	926	100
1301	gi17382046	Homo sapiens	unnamed protein product	926	100
1301	gi2662057	Rattus norvegicus	RGS8	921	98
1302	AAB95302	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17538	1242	100
1302	gi10434969	Homo sapiens	cDNA FLJ13105 fis, clone NT12RP3002351, weakly similar to Human mRNA for NAD-dependent methylene tetrahydrofolate dehydrogenase cyclohydrolase (EC 1.5.1.15).	1242	100
1302	gi12834726	Mus musculus	putative	873	94
1303	gi546603	human, tumorous liver, mRNA Partial, 2631 nt]. [Homo sapiens	glutamine synthetase	1787	100
1303	gi17512038	Homo sapiens	clone MGC:20095 IMAGE:3352740, mRNA, complete cds.	1787	100
1303	gi15080157	Homo sapiens	glutamate-ammonia ligase (glutamine synthase), clone MGC:20322 IMAGE:4137547, mRNA, complete cds.	1787	100
1304	gi546603	human, tumorous liver, mRNA Partial, 2631 nt]. [Homo sapiens	glutamine synthetase	432	89
1304	gi31833	Homo sapiens	Human mRNA for glutamine synthetase (E.C. 6.3.1.2).	432	89
1304	gi31831	Homo sapiens	Human rearranged mRNA for glutamine synthase.	432	89
1305	gi16551755	Homo sapiens	cDNA FLJ131807 fis, clone NT12RI2009215, moderately similar to ZINC FINGER	492	54

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			PROTEIN 165.		
1305	AAM41649	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6580.	462	51
1305	AAM93917	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 4075.	462	51
1306	AAU16246	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1199.	1010	96
1306	gi12832845	Mus musculus	putative	585	83
1306	AAU16240	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1193.	342	95
1307	gi9798452	Homo sapiens	mRNA for putative capacitative calcium channel (trp7 gene).	4470	100
1307	gi5326854	Mus musculus	receptor-activated calcium channel	4392	98
1307	gi2295903	Homo sapiens	Human putative calcium influx channel (htrp3) mRNA, complete cds.	3529	81
1308	gi6114601	Homo sapiens	mRNA for stromal antigen 3 (STAG3 gene).	281	74
1308	gi3090423	Mus musculus	stag3	203	49
1308	gi13195163	Rattus norvegicus	stromal antigen 3	199	47
1309	gi6114601	Homo sapiens	mRNA for stromal antigen 3 (STAG3 gene).	295	82
1309	gi3090423	Mus musculus	stag3	200	55
1309	gi13195163	Rattus norvegicus	stromal antigen 3	198	54
1310	gi9858562	Homo sapiens	Rh type B glycoprotein (RHBG) mRNA, complete cds.	2176	99
1310	gi15718471	Homo sapiens	Rh type B glycoprotein (RHBG) gene, exons 9, and 10 and complete cds.	2176	99
1310	gi14346006	Pan troglodytes	Rh type B glycoprotein	2161	99
1311	gi7243149	Homo sapiens	mRNA for KIAA1384 protein, partial cds.	3377	100
1311	gi12857673	Mus musculus	putative	2817	98
1311	gi7242973	Homo sapiens	mRNA for KIAA1309 protein, partial cds.	913	33
1312	AAY86297	Homo sapiens	HUMA- Human secreted protein HLDCE79, SEQ ID NO:212.	530	100
1312	AAY21623	Homo sapiens	REGC Ligand binding domain of nuclear receptor hGR.	74	32
1312	AAP80919	Homo sapiens	SALK Sequence of the primary protein sequence of human glucocorticoid receptor (hGR).	74	32
1313	gi11022690	Homo sapiens	ifp1 mRNA for interferon-responsive finger protein 1 long form, complete cds.	4302	99
1313	AAB95586	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18252.	2612	100
1313	gi14042869	Homo sapiens	cDNA FLJ14970 fis, clone THYR01000501, weakly	2612	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			similar to 52 KD RO PROTEIN.		
1314	gi17861372	Homo sapiens	lysyl oxidase-like 4 mRNA, complete cds.	984	93
1314	gi16660131	Homo sapiens	lysyl oxidase-like 4 (LOXL4) mRNA, complete cds.	984	93
1314	gi14669471	Homo sapiens	lysyl oxidase-related protein C (LOXC) mRNA, complete cds.	984	93
1315	gi16549449	Homo sapiens	cDNA FLJ30273 fis, clone BRACE2002685, moderately similar to Homo sapiens androgen-regulated short-chain dehydrogenase/reductase I (ARSDRI) mRNA.	1612	98
1315	gi12861668	Mus musculus	putative	1374	84
1315	gi16740649	Mus musculus	Similar to RIKEN cDNA A930033N07 gene	1315	81
1316	gi14286286	Homo sapiens	Similar to hypothetical protein FLJ20515, clone MGC:2696 IMAGE2820596, mRNA, complete cds.	1006	100
1316	AAV53023	Homo sapiens	GEMY Human secreted protein clone q662.3 protein sequence SEQ ID NO:52.	990	99
1316	AAE04835	Homo sapiens	SUGE-Human SGP001 phosphatase polypeptide.	931	95
1317	AAN50069_aa1	Homo sapiens	MITU DNA encoding cardiolipin in plasmid pHANF48.	771	100
1317	AAW98193	Homo sapiens	CURA- Human atrial natriuretic peptide prohormone.	771	100
1317	AAP51241	Homo sapiens	BIOT- Sequence of pre-pro-atrial natriuretic/vasodilator polypeptide (ANVP).	771	100
1318	AAE06183	Homo sapiens	HUMA- Human gene 57 encoded secreted protein fragment, SEQ ID NO:245.	3182	89
1318	AAV87206	Homo sapiens	HUMA- Human secreted protein sequence SEQ ID NO:245.	3182	89
1318	AAE06097	Homo sapiens	HUMA- Human gene 57 encoded secreted protein HRAC80, SEQ ID NO:159.	2906	88
1319	AAU08995	Homo sapiens	MILL- Human G protein-coupled receptor, GPCR, 45449.	410	96
1319	gi12214287	Homo sapiens	Human DNA sequence from clone RP3-402H5 on chromosome 6p12.3-21.1. Contains ESTs, STSs and GSSs. Contains the 3' part of a gene for a novel 7 transmembrane receptor of the rhodopsin family and a novel gene, complete sequence.	410	96
1319	gi15797318	Homo sapiens	unnamed protein product	410	96
1320	gi897761	Homo sapiens	H.sapiens mRNA for protein	387	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1320	gi4558638	Homo sapiens	phosphatase 5, chromosome 19, BAC 82621 (CIT-B-139a18), complete sequence.	387	100
1320	gi12805033	Homo sapiens	protein phosphatase 5, catalytic subunit, clone MGC:5260 IMAGE:3459309, mRNA, complete cds.	387	100
1321	AAB56613	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1191.	1174	100
1321	gi182644	Homo sapiens	Human FK506-binding protein 25 (FKBP25) mRNA, complete cds.	1169	100
1321	gi182626	Homo sapiens	Human rapamycin binding protein (FK506) mRNA, complete cds.	1169	100
1322	gi15042691	Homo sapiens	sorting nexin 18 (SNX18) mRNA, complete cds.	2895	100
1322	gi15559064	Mus musculus	SNAG1	2440	86
1322	AAW99023	Homo sapiens	MOUN 17G2 peptide sequence.	1605	95
1323	gi12804803	Homo sapiens	clone MGC:4499 IMAGE:2964565, mRNA, complete cds.	1266	100
1323	gi12654515	Homo sapiens	clone MGC:2827 IMAGE:2964565, mRNA, complete cds.	1266	100
1323	AAB54374	Homo sapiens	HUMA- Human pancreatic cancer antigen protein sequence SEQ ID NO:826.	1261	99
1324	gi11907599	Homo sapiens	protein kinase HIPK2 mRNA, complete cds.	6242	99
1324	AAB65661	Homo sapiens	SUGE- Novel protein kinase, SEQ ID NO: 188.	6086	97
1324	gi17225377	Homo sapiens	homeodomain interacting protein kinase 2 (HIPK2) mRNA, complete cds.	6083	97
1325	AAB65661	Homo sapiens	SUGE- Novel protein kinase, SEQ ID NO: 188.	6124	99
1325	gi17225377	Homo sapiens	homeodomain interacting protein kinase 2 (HIPK2) mRNA, complete cds.	6121	99
1325	gi11907599	Homo sapiens	protein kinase HIPK2 mRNA, complete cds.	6072	97
1326	gi16552298	Homo sapiens	cDNA FLJ32230 fis, clone PLACE6004464, weakly similar to Human placenta (Diff48) mRNA.	3064	99
1326	gi13274202	Homo sapiens	Human DNA sequence from clone RP4-530115 on chromosome 20 Contains the 3' end of the PTPN1 gene for protein tyrosine phosphatase, non-receptor type 1 (EC 3.1.3.48), the gene for a novel	2261	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			protein similar to placental protein DIFF40, an RPL36 (60S ribosomal protein L36) pseudogene, a novel gene, two putative novel genes, ESTs, STSs and GSSs, complete sequence.		
1326	gi2224713	Homo sapiens	Human mRNA for KIAA0386 gene, complete cds.	963	34
1327	AAU12177	Homo sapiens	GETH Human PRO305 polypeptide sequence.	1472	82
1327	AAV81487	Homo sapiens	FUJY Human cathepsin L2.	1472	82
1327	AAV02358	Homo sapiens	ONOV Polypeptide identified by the signal sequence trap method.	1472	82
1328	AAU12177	Homo sapiens	GETH Human PRO305 polypeptide sequence.	1698	84
1328	AAV81487	Homo sapiens	FUJY Human cathepsin L2.	1698	84
1328	AAV02358	Homo sapiens	ONOV Polypeptide identified by the signal sequence trap method.	1698	84
1329	AAV87329	Homo sapiens	INCY- Human signal peptide containing protein HSPP-106 SEQ ID NO:106.	692	94
1329	gi15145428	Caenorhabditis elegans	Hypothetical protein Y22D7AL.14	74	23
1329	gi3618016	Human immunodeficiency virus type 1	nef	73	35
1330	AAB82315	Homo sapiens	UYCO Human immunoglobulin receptor isoform IRTA2c.	1120	99
1330	AAB82314	Homo sapiens	UYCO Human immunoglobulin receptor isoform IRTA2b.	1120	99
1330	AAB82313	Homo sapiens	UYCO Human immunoglobulin receptor isoform IRTA2a.	1120	99
1331	gi14336757	Homo sapiens	16p13.3 sequence section 6 of 8.	1178	100
1331	gi13436269	Homo sapiens	hypothetical protein FLJ20898, clone MGC:10688 IMAGE:3622114, mRNA, complete cds.	1178	100
1331	AAG81430	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:378.	954	100
1332	gi5689557	Homo sapiens	mRNA for KIAA1110 protein, partial cds.	3820	99
1332	ABB11713	Homo sapiens	HYSE- Human KIAA1110 protein homologue, SEQ ID NO:2083.	3809	99
1332	gi3882247	Homo sapiens	mRNA for KIAA0763 protein, complete cds.	1587	44
1333	gi13366277	Homo sapiens	Human DNA sequence from clone RP5-998H6 on chromosome 20q13.1. Contains the gene for the ortholog of rat PB-Cadherin, ESTs, STSs, GSS, two CpG islands and genomic marker D20S17, complete	4283	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			sequence.		
1333	gi4760578	Mus musculus	PB-Cadherin	3918	92
1333	gi1398906	Rattus norvegicus	long type PB-cadherin	3907	92
1334	gi3023044	Enterococcus faecalis	orfC	302	45
1334	gi496520	Streptococcus pyogenes	orf iota	271	39
1334	gi12957024	Enterococcus faecalis	hypothetical protein	251	38
1335	gi12838481	Mus musculus	putative	491	71
1335	gi2072497	Mus musculus	perforatorial protein PERF 15	481	68
1335	gi1519523	Rattus norvegicus	PERF15 protein	477	68
1336	gi10438454	Homo sapiens	cDNA: FLJ22171 fis, clone HRC00654.	3750	100
1336	gi10438150	Homo sapiens	cDNA: FLJ21935 fis, clone HEP04373.	3734	99
1336	gi1352954	Mus musculus	Similar to hypothetical protein FLJ21935	3203	85
1337	gi12803445	Homo sapiens	clone MGC:2217 IMAGE:3139026, mRNA, complete cds.	463	100
1337	gi7296344	Drosophila melanogaster	CG4186 gene product	201	47
1337	gi529133	Saccharomyces cerevisiae	Yhr116wp	113	41
1338	gi10434352	Homo sapiens	cDNA FLJ12697 fis, clone NT2RP1000522, weakly similar to UBIQUITIN CARBOXYL-TERMINAL HYDROLASE DUB-1 (EC 3.1.2.15).	6400	99
1338	AAB95146	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17169.	6396	99
1338	AAB74671	Homo sapiens	INCY- Human protease and protease inhibitor PPIM-4.	4021	99
1339	gi10801626	Macaca fascicularis	hypothetical protein	1668	98
1339	gi12836718	Mus musculus	putative	1439	84
1339	gi10438311	Homo sapiens	cDNA: FLJ22054 fis, clone HEP09634.	1351	99
1340	gi17512067	Homo sapiens	hypothetical protein DKFZp434D0421, clone MGC:20807 IMAGE:4330507, mRNA, complete cds.	1903	100
1340	gi14043717	Homo sapiens	hypothetical protein DKFZp434D0421, clone MGC:14446 IMAGE:4304040, mRNA, complete cds.	1903	100
1340	gi12053119	Homo sapiens	mRNA; cDNA DKFZp434D0421 (from clone DKFZp434D0421); complete cds.	1903	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1341	AAB08765	Homo sapiens	INCY- A human leukocyte and blood related protein (LBAP).	716	93
1341	AAM40991	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5922.	508	93
1341	AAB74718	Homo sapiens	INCY- Human membrane associated protein MEMAP-24.	456	96
1342	AAB95563	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18202.	3214	99
1342	gi10435904	Homo sapiens	cDNA FLJ13782 f1s, clone PLACE4000489, weakly similar to PROTEIN GRAINY-HEAD.	3214	99
1342	gi12832762	Mus musculus	putative	2094	94
1343	AAG89336	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 456.	801	100
1343	AAG81352	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:222.	801	100
1343	AAY91423	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 11 SEQ ID NO:144.	801	100
1344	AAY94978	Homo sapiens	QEMY Human secreted protein clone pw337_6 2nd protein sequence SEQ ID NO:238.	444	100
1344	gi7594534	Arabidopsis thaliana	putative protein	79	31
1345	gi4309894	Homo sapiens	PAC clone RP4-555L14 from 7q34-q36, complete sequence.	818	100
1345	gi17646448	Mus musculus	gammaN-crystallin	724	83
1345	gi17646446	Homo sapiens	gammaN-crystallin variant (CRYGN) mRNA, complete cds.	600	100
1346	AAY76216	Homo sapiens	HUMA- Human secreted protein encoded by gene 93.	225	97
1347	AAY11447	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID No 269.	210	97
1347	gi7619769	Streptomyces coelicolor A3(2)	probable LacI-family transcriptional regulatory protein.	66	53
1347	gi755580	Streptomyces lividans	ORF-RDR; LacI homolog, similar to E. coli Lac repressor, Swiss-Prot Accession Number P03023	66	53
1348	gi17391052	Homo sapiens	clone MGC:9915 IMAGE:3871205, mRNA, complete cds.	2220	100
1348	AAG74153	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4917.	1072	98
1348	gi525318	Haematobia irritans	putative ATPase	937	44
1349	gi17391052	Homo sapiens	clone MGC:9915 IMAGE:3871205, mRNA, complete cds.	1919	88
1349	AAG74153	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4917.	782	76
1349	gi525318	Haematobia	putative ATPase	749	39

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		irritans			
1350	AAB43315	Homo sapiens	CURA- Human ORFX ORF3079 polypeptide sequence SEQ ID NO:6158.	1463	99
1350	gi6682590	Homo sapiens	chromosome 14 clones RP11-111016 and RP11-61F4 containing genes for nuclear receptor coactivator NCoA-62 (nuclear receptor coactivator NCoA-62) gene, complete cds; and unknown gene, complete sequence.	1463	99
1350	AAW8203	Homo sapiens	GEMY Human foetal brain secreted protein fh3_6 (alternative sequence).	1249	99
1351	gi12841145	Mus musculus	putative	1153	93
1351	gi13529212	Homo sapiens	Similar to RIKEN cDNA 1810018M11 gene, clone MGC:12485 IMAGE:3932127, mRNA, complete cds.	1136	99
1351	AAV96202	Homo sapiens	UYNY IkappaB kinase (IKK) binding protein, Y2H56.	1126	98
1352	gi15126788	Mus musculus	Similar to ferritin heavy chain	947	100
1352	gi50954	Mus musculus	ferritin H subunit	947	100
1352	gi50952	Mus musculus	ferritin heavy subunit (AA 1 - 182)	947	100
1353	AAB70538	Homo sapiens	CURA- Human PRO8 protein sequence SEQ ID NO:16.	2777	98
1353	AAB70537	Homo sapiens	CURA- Human PRO7 protein sequence SEQ ID NO:14.	2777	98
1353	gi13185725	Homo sapiens	n 1755 can be A, G, C, or T.	2777	98
1354	AAB53541	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:1081.	110	73
1354	AAR72986	Homo sapiens	GENZ Creatine-kinase subunit B.	110	73
1354	gi29963	Homo sapiens	Human gene for creatine kinase B (EC 2.7.3.2).	110	73
1355	gi12407405	Homo sapiens	tripartite motif protein TRIM9 isoform beta (TRIM9) mRNA, complete cds; alternatively spliced.	2831	100
1355	gi16755524	Rattus norvegicus	Spring	2783	97
1355	gi1665803	Homo sapiens	Human mRNA for KIAA0282 gene, partial cds.	2575	99
1356	AAV53641	Homo sapiens	CHIR A bone marrow secreted protein designated BMS42.	346	98
1356	gi9663153	Homo sapiens	partial mRNA for transport-secretion protein 2.2, (TTS-2.2 gene).	346	98
1356	gi9663151	Homo sapiens	partial mRNA for transport-secretion protein 2.1 (TTS-2.1 gene).	346	98

SEQ ID	Hlt ID	Species	Description	S score	Percent identity
1357	gi10801197	Homo sapiens	heparanase-like protein HPA2b mRNA, complete cds.	2785	100
1357	gi15132669	Homo sapiens	unnamed protein product	2785	100
1357	AAA91097_aa1	Homo sapiens	INSI- Human heparanase, hnhp1, coding sequence.	2626	88
1358	gi6330358	Homo sapiens	mRNA for KIAA1193 protein, partial cds.	2885	100
1358	AAU16216	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1169.	1216	88
1358	AAV15835	Homo sapiens	PATE/ A human er1 protein.	751	43
1359	AAB74945	Homo sapiens	YAMA Human ADAM type metal protease MDT2S protein SEQ ID NO:10.	6065	99
1359	gi11493589	Homo sapiens	zinc metalloendopeptidase (ADAMTS10) mRNA, partial cds.	5940	99
1359	AAB72300	Homo sapiens	HIRO/ Human ADAMTS-10 alternative amino acid sequence.	5484	97
1360	gi17384256	Homo sapiens	partial MUC5AC gene for mucin 5, clone A.	1291	80
1360	gi563375	Homo sapiens	H.sapiens (JER47) MUC5AC mRNA for mucin (partial).	978	91
1360	gi17384254	Homo sapiens	partial mRNA for mucin 5 (MUC5AC gene).	905	75
1361	AAM24395	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1920.	634	100
1361	AAV76179	Homo sapiens	HUMA- Human secreted protein encoded by gene 56.	634	100
1361	AAB47327	Homo sapiens	CURA- FCTR4.	74	27
1362	gi15080264	Homo sapiens	clone MGC:20279 IMAGE:3949150, mRNA, complete cds.	1043	100
1362	gi10439083	Homo sapiens	cDNA: FLJ22623 fls, clone HSI05687.	1043	100
1362	gi17389437	Homo sapiens	hypothetical protein FLJ22623, clone MGC:22173 IMAGE:4274089, mRNA, complete cds.	1031	99
1363	AAH78730_aa1	Homo sapiens	HUMA- Human HIBCJ89 serine/threonine phosphatase cDNA sequence.	1635	99
1363	AAU20555	Homo sapiens	HUMA- Human secreted protein, Seq ID No 547.	1635	99
1363	AAU20663	Homo sapiens	HUMA- Human secreted protein, Seq ID No 655.	1635	99
1364	AAB95700	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18526.	752	100
1364	gi15341846	Homo sapiens	hypothetical protein FLJ14107, clone MGC:21030 IMAGE:4389733, mRNA, complete cds.	752	100
1364	gi10436485	Homo sapiens	cDNA FLJ14107 fls, clone MAMMA1001252.	752	100
1365	AAV673	Homo sapiens	GEMY Human secreted protein	367	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	16		BL341_4 amino acid sequence.		
1365	AAAY08625	Homo sapiens	GEMY Human secreted protein BL341_4.	367	100
1365	AAW42018	Homo sapiens	JACO/Clone BL341_4 protein.	363	98
1366	AAU19715	Homo sapiens	HUMA- Human novel extracellular matrix protein, Seq ID No 365.	2225	99
1366	gi6136798	Mus musculus	synaptotagmin VIdeltaTM2	2150	96
1366	gi6136796	Mus musculus	synaptotagmin VIdeltaTM1	2150	96
1367	AAU19715	Homo sapiens	HUMA- Human novel extracellular matrix protein, Seq ID No 365.	2173	97
1367	gi6136798	Mus musculus	synaptotagmin VIdeltaTM2	2098	94
1367	gi6136796	Mus musculus	synaptotagmin VIdeltaTM1	2098	94
1368	AAE05175	Homo sapiens	INCY- Human drug metabolising enzyme (DME-6) protein.	2614	97
1368	AAU12225	Homo sapiens	GETH Human PRO4404 polypeptide sequence.	2614	97
1368	gil1933056	Sus scrofa	cytochrome P450	1305	50
1369	AAW78135	Homo sapiens	HUMA- Human secreted protein encoded by gene 10 clone HPMGQ80.	385	100
1369	AAO02310	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 16202.	76	39
1369	AAO08772	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 22664.	74	35
1370	gi10047301	Homo sapiens	mRNA for KIAA1613 protein, partial cds.	3532	100
1370	AAB36171	Homo sapiens	LEXI- Novel human transporter protein SEQ ID NO: 28.	3412	100
1370	AAB36170	Homo sapiens	LEXI- Novel human transporter protein SEQ ID NO: 26.	3408	99
1371	AAB41673	Homo sapiens	CURA- Human ORFX ORF1437 polypeptide sequence SEQ ID NO:2874.	1221	96
1371	AAB61611	Homo sapiens	PROT- Human protein HP03377.	1220	100
1371	AAE03656	Homo sapiens	INCY- Human extracellular matrix and cell adhesion molecule-20 (XCAD-20).	1220	100
1372	gi5295882	Mus musculus	kinesin like protein 9	3618	88
1372	AAB94768	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15849.	3257	99
1372	gi10435968	Homo sapiens	cDNA FLJ13832 fis, clone THYRO1000666, highly similar to Mus musculus mRNA for kinesin like protein 9.	3257	99
1373	AAH25568_aal	Homo sapiens	CURA- Nucleotide sequence of an interferon omega-1 like	3294	100

SEQ ID	HIT ID	Species	Description	S score	Percent identity
			protein NOV2.		
1373	AAG67523	Homo sapiens	SMIK Amino acid sequence of a human secreted polypeptide.	3294	100
1373	AAB84469	Homo sapiens	CURA- Amino acid sequence of an interferon omega-1 like protein NOV2.	3294	100
1374	gi5923992	Homo sapiens	Human DNA sequence from clone RP5-1043E3 on chromosome 6p21.1-21.2. Contains part of a novel gene, an transcription factor E2F4 pseudogene, ESTs, STSs and GSSs, complete sequence.	945	99
1374	gi14245713	Giardia intestinalis	kinesin-like protein 9	543	44
1374	gi15022394	Leishmania major	possible kinesin-like protein	531	42
1375	AAG74779	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5543.	1405	99
1375	gi7981261	Homo sapiens	Human DNA sequence from clone RP1-50024 on chromosome 1p35.1-35.3. Contains the 3' end of the gene for a novel protein similar to C. elegans K07B1.7 (Tr:O01886), the gene for a novel protein (translation of cDNA NTZRM2001100 (Em:AK001211)), the SFN gene for stratifin (14-3-3 protein sigma), the gene for a novel protein with DHHC zinc finger domain, the gene for a novel protein (translation of cDNA KAT07271 (Em:AK000484)) and the gene for B120 (C1orf4) (ARID DNA binding domain containing protein). Contains ESTs, STSs, GSSs and six putative CpG islands, complete sequence.	1122	67
1375	AAG81254	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:26.	1118	72
1376	gi8052237	Homo sapiens	ASCL3 gene, CEGP1 gene, C11orf14 gene, C11orf15 gene, C11orf16 gene and C11orf17 gene.	5605	100
1376	gi8052320	Mus musculus	Cegp1 protein	5054	89
1376	AAG67529	Homo sapiens	SMIK Amino acid sequence of a human secreted polypeptide.	3226	61
1377	gi10437045	Homo sapiens	cDNA: FLJ21044 fis, clone CAE11659.	1663	100
1377	gi4206386	Mus musculus	rig-1 protein	1543	72
1377	AAB57088	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ	1518	98

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			ID NO:1666.		
1378	AAY02697	Homo sapiens	HUMA- Human secreted protein encoded by gene 48 clone HTNBR95.	165	100
1379	AAY73386	Homo sapiens	INCY- HTRM clone 3279329 protein sequence.	529	100
1379	AAB63162	Homo sapiens	ROSE/ Human secreted protein sequence encoded by gene 29 SEQ ID NO:88.	363	100
1379	AAB95124	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17122.	83	32
1380	gi10198115	Homo sapiens	2P domain potassium channel TREK2 (KCNK10) mRNA, complete cds.	2760	100
1380	gi8452900	Rattus norvegicus	potassium channel TREK-2	2555	95
1380	gi4584799	Mus musculus	TREK-1 K+ channel subunit	1238	64
1381	gi13276655	Homo sapiens	mRNA; cDNA DKFZp761M0423 (from clone DKFZp761M0423); complete cds.	2626	99
1381	AAE04361	Homo sapiens	INCY- Human kinase (PKIN)-2.	2588	97
1381	gi1836161	Rattus sp.	Ca2+/calmodulin-dependent protein kinase IV kinase isoform; CaM-kinase kinase alpha	2468	93
1382	gi12382781	Homo sapiens	OSBP-related protein 4 mRNA, complete cds.	1124	100
1382	gi133359201	Homo sapiens	mRNA for KIAA1664 protein, partial cds.	1036	100
1382	gi14209840	Homo sapiens	oxysterol binding protein 2 (OSBP2) gene, complete cds.	919	100
1383	gi12805533	Mus musculus	Unknown (protein for MGC:7583)	792	99
1383	gi12858656	Mus musculus	putative	787	98
1383	AAM23865	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1390.	672	83
1384	gi3002925	Homo sapiens	T cell receptor beta chain (TCRBV13S1-TCRBJ2S1) mRNA, complete cds.	679	73
1384	gi2982508	Homo sapiens	mRNA for TCR beta chain, specific for Mage 3/HLA-A2.	667	71
1384	AAM24051	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1576.	655	100
1385	gi13879262	Mus musculus	RIKEN cDNA 0610011E17 gene	710	97
1385	gi12850353	Mus musculus	putative	710	97
1385	AAB42905	Homo sapiens	CURA- Human ORFX ORF2669 polypeptide sequence SEQ ID NO:5338.	582	79
1386	AAB95062	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16896.	741	99
1386	AAM241	Homo sapiens	HYSE- Human EST encoded	681	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	27		protein SEQ ID NO: 1652.		
1386	AAAY87328	Homo sapiens	INCY- Human signal peptide containing protein HSPP-105 SEQ ID NO:105.	681	100
1387	AAM23576	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1101.	691	100
1387	AAB71899	Homo sapiens	ZYMO Human alpha48.	306	68
1387	AAE06580	Homo sapiens	SAGA Human protein having hydrophobic domain, HP10786.	306	68
1388	AAB71863	Homo sapiens	MILL- Human h15571 GPCR.	6511	97
1388	gi15987491	Homo sapiens	tumor endothelial marker 5 precursor (TEM5) mRNA, complete cds.	6511	97
1388	gi15987499	Mus musculus	tumor endothelial marker 5 precursor	5693	85
1389	AAE01354	Homo sapiens	HUMA- Human gene 3 encoded secreted protein HOHBL42, SEQ ID NO:76.	3747	99
1389	gi431420	Mus musculus	MPS1 protein	2714	77
1389	gi5052048	Rattus norvegicus	Mpg-1 protein	2672	75
1390	AAM24200	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1725.	329	100
1390	AAAY19588	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	329	100
1391	AAG74249	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5013.	888	98
1391	gi12835118	Mus musculus	putative	884	95
1391	gi7301763	Drosophila melanogaster	CG11900 gene product	497	59
1392	gi12803269	Homo sapiens	Similar to CG10641 gene product, clone MGC:3052 IMAGE:343900, mRNA, complete cds.	701	100
1392	gi10441942	Homo sapiens	clone PP3051 unknown mRNA.	701	100
1392	AAAB95496	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18039.	698	99
1393	AAAY12512	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO:543.	320	98
1394	AAAB94833	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15998.	2735	100
1394	gi10436242	Homo sapiens	cDNA FLJ13941 fis, clone Y79AA1000850.	2735	100
1394	AAAB42818	Homo sapiens	CURA- Human ORFX ORF2582 polypeptide sequence SEQ ID NO:5164.	1115	99
1395	AAAB94833	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15998.	2713	97
1395	gi10436242	Homo sapiens	cDNA FLJ13941 fis, clone Y79AA1000850.	2713	97
1395	AAAB42818	Homo sapiens	CURA- Human ORFX ORF2582 polypeptide sequence	1093	94

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			SEQ ID NO:5164.		
1396	gi7263928	Homo sapiens	Human DNA sequence from clone RP1-61A9 on chromosome 1p35.2-36.13 Contains part of the EPHB2 gene for tyrosine-protein kinase, the gene KIAA0478 for a C2H2 type zinc finger gene, BSTs, STSs, GSSs and three putative CpG Islands, complete sequence.	5326	99
1396	AAE04362	Homo sapiens	INCY- Human kinase (PKIN)-3.	5308	99
1396	AAU00691	Homo sapiens	CURA- Ephrin type-A receptor 8-like protein.	5259	99
1397	gi10437626	Homo sapiens	cDNA: FLJ21511 fis, clone COL05748.	3713	99
1397	gi16741400	Mus musculus	Similar to hypothetical protein FLJ21511	3125	82
1397	gi14588931	Saccharomyces cerevisiae	hypothetical protein	690	29
1398	AAB94713	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15714.	2869	100
1398	gi10435745	Homo sapiens	cDNA FLJ13664 fis, clone PLACB1011649.	2869	100
1398	gi16877291	Homo sapiens	Similar to hypothetical protein 24432, clone MGC:21034 IMAGE:4400396, mRNA, complete cds.	2843	99
1399	AAB82940	Homo sapiens	UYNY Human androgen receptor trapped protein 5 (ART5).	1429	100
1399	AAB56085	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 9 SEQ ID NO:179.	1429	100
1399	gi10439204	Homo sapiens	cDNA: FLJ22709 fis, clone HSI13338.	1429	100
1400	gi10439625	Homo sapiens	cDNA: FLJ23056 fis, clone LNG03287.	1190	100
1400	gi7295732	Drosophila melanogaster	ft gene product	185	27
1400	gi157409	Drosophila melanogaster	fat protein	185	27
1401	gi13936285	Mus musculus	TRH4	1332	61
1401	gi12845540	Mus musculus	putative	1330	61
1401	AAU00782	Homo sapiens	INCY- Human apoptosis protein, APOP-2.	1092	65
1402	AAU00475	Homo sapiens	MILL- Human INTERCEPT 394 alternative form protein.	4272	97
1402	AAU00473	Homo sapiens	MILL- Human INTERCEPT 394 protein.	4089	99
1402	gi10438450	Homo sapiens	cDNA: FLJ22169 fis, clone HRC00632.	3505	99
1403	gi10798804	Homo sapiens	HCMOGT-1 mRNA for sperm antigen, complete cds.	3737	98
1403	ABB1229	Homo sapiens	HYSE- Human secreted protein	2811	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	7		homologue, SEQ ID NO:2667.		
1403	AAM25255	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:770.	2778	98
1404	AAB95425	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17833.	4871	100
1404	gi10435487	Homo sapiens	cDNA FLJ13465 fis, clone PLACE1003493, weakly similar to ENDOTHELIAL CELL MULTIMERIN PRECURSOR.	4871	100
1404	AA Y36300	Homo sapiens	HUMA- Human secreted protein encoded by gene 77.	2472	98
1405	gi10047211	Homo sapiens	mRNA for KIAA1573 protein, partial cds.	6270	100
1405	gi1438834	Macaca fascicularis	hypothetical protein	5174	99
1405	AAB95883	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18994.	3679	100
1406	gi14714604	Homo sapiens	clone MGC:17248 IMAGE:4215164, mRNA, complete cds.	3291	96
1406	AAW80318	Homo sapiens	SMIK Neurodegenerative polypeptide HHPDZ65var.	2893	100
1406	gi8346834	Homo sapiens	mRNA for putative acid-sensing ion channel (ASIC4 gene).	2893	100
1407	AAB65697	Homo sapiens	SUGE- Novel protein kinase, SEQ ID NO: 225.	1647	100
1407	gi14043928	Homo sapiens	clone IMAGE:4139786, mRNA, partial cds.	1117	100
1407	AAG02479	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6560.	442	100
1408	AAR49943	Homo sapiens	YAMA/ Human hippocampal cholinergic neurotrophic peptide precursor.	881	89
1408	gi704465	Homo sapiens	H.sapiens mRNA for phosphatidylethanolamine binding protein.	881	89
1408	gi435638	Homo sapiens	Human mRNA for human homologue of rat phosphatidylethanolamine binding protein, complete cds.	881	89
1409	AAB95517	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18089.	4520	99
1409	gi10435776	Homo sapiens	cDNA FLJ13687 fis, clone PLACE2000061.	4520	99
1409	gi7264653	Mus musculus	Kiaa0575	1867	48
1410	gi17028341	Homo sapiens	hypothetical protein FLJ21820, clone MGC:14932 IMAGE:3611020, mRNA, complete cds.	1732	100
1410	gi10437997	Homo sapiens	cDNA: FLJ21820 fis, clone HEP01232.	1732	100
1410	gi16769718	Drosophila melanogaster	LP01162p	437	33
1411	gi15079729	Homo sapiens	hypothetical protein FLJ21125, clone MGC:14948 IMAGE:4303449, mRNA,	1530	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			complete cds.		
1411	gi10437158	Homo sapiens	cDNA: FLJ21125 fis, clone CAS06077.	1530	99
1411	AAV94674	Homo sapiens	ZYMO Human zsig83 mature protein sequence.	142	34
1412	gi10438126	Homo sapiens	cDNA: FLJ21918 fis, clone HEP04006.	3799	100
1412	AAV84440	Homo sapiens	INCY- Amino acid sequence of a human RNA-associated protein.	2085	59
1412	gi7020094	Homo sapiens	cDNA FLJ20171 fis, clone COL09761.	1246	66
1413	gi14031072	Homo sapiens	Human DNA sequence from clone RP3-331H24 on chromosome 6 Contains a putative novel gene, part of the gene for hypothetical protein FLJ21079, similar to opioid growth factor receptor, ESTs, STSs, GSSs and a CpG island, complete sequence.	1307	99
1413	gi10437094	Homo sapiens	cDNA: FLJ21079 fis, clone CAS02253.	1307	99
1413	gi12843468	Mus musculus	putative	922	73
1414	AAB943	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14968.	3638	100
1414	gi10434785	Homo sapiens	cDNA FLJ12987 fis, clone NT2RP3000068, weakly similar to SON OF SEVENLESS PROTEIN HOMOLOG 1.	3638	100
1414	AAB95639	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18376.	3001	100
1415	gi13623491	Homo sapiens	clone MGC:13125 IMAGE:4111572, mRNA, complete cds.	3054	100
1415	gi16553793	Homo sapiens	cDNA FLJ25103 fis, clone CBR01405.	1586	74
1415	AAM66279	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26585.	1301	100
1416	ABB11675	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2045.	1238	99
1416	gi15928921	Homo sapiens	hypothetical protein FLJ14393, clone MGC:17935 IMAGE:3916148, mRNA, complete cds.	1238	99
1416	AAV45272	Homo sapiens	HUMA- Human secreted protein encoded from gene 16.	1236	99
1417	gi13623249	Homo sapiens	Similar to RIKEN cDNA 311082I17 gene, clone MGC:11257 IMAGE:3941780, mRNA, complete cds.	945	95
1417	gi12852007	Mus musculus	putative	466	61
1417	AAW67936	Homo sapiens	HUMA- Fragment of human secreted protein encoded by	329	98

SEQ ID	Hit ID	Species	Description	S score	Percent Identity
			gene 57.		
1418	gi13938274	Homo sapiens	clone MGC:15548 IMAGE:3051320, mRNA, complete cds.	3136	99
1418	AAU16373	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1326.	963	100
1418	AAU15922	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 875.	963	100
1419	AAB93081	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11914.	1812	98
1419	gi14042419	Homo sapiens	cDNA FLJ14712 fis, clone NT2RP3000825, weakly similar to NEUROGENIC LOCUS NOTCH 3 PROTEIN.	1812	98
1419	AAU72713	Homo sapiens	HUMA- HWAQA40 clone human attractin-like protein.	1212	99
1420	AAU12418	Homo sapiens	GETH Human PRO1275 polypeptide sequence.	643	98
1420	AAU99379	Homo sapiens	GETH Human PRO1275 (UNQ645) amino acid sequence SEQ ID NO:136.	643	98
1420	AAB25683	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 19 SEQ ID NO:72.	643	98
1421	gi10438712	Homo sapiens	cDNA: FLJ22358 fis, clone HRC06415.	3025	100
1421	gi14211139	Homo sapiens	NADPH oxidase 5 gamma mRNA, complete cds.	3019	99
1421	gi14211137	Homo sapiens	NADPH oxidase 5 alpha mRNA, complete cds.	3019	99
1422	gi12658305	Homo sapiens	kappa B and V(D)J recombination signal sequences binding protein (KRC) mRNA, complete cds.	8934	99
1422	gi10047175	Homo sapiens	mRNA for KIAA1555 protein, partial cds.	8588	99
1422	gi1377886	Mus musculus	DNA binding protein Rc	6216	76
1423	gi17389208	Homo sapiens	clone MGC:16889 IMAGE:3883022, mRNA, complete cds.	2465	100
1423	gi15278167	Homo sapiens	differentiation-related DIF14 long form (DIF14) mRNA, complete cds, alternatively spliced.	2448	99
1423	gi9651220	Mus musculus	LMBR1 long form	2391	96
1424	gi10437078	Homo sapiens	cDNA: FLJ21069 fis, clone CAS01594.	2523	99
1424	gi15929778	Homo sapiens	hypothetical protein FLJ21069, clone MGC:21026 IMAGE:4431888, mRNA, complete cds.	2517	99
1424	gi1285974	Mus musculus	putative	2182	86
1425	AAM93735	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3701.	1364	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1425	gi12855379	Mus musculus	putative	1332	87
1425	gi16359363	Mus musculus	Similar to hypothetical protein MGC12921	652	53
1426	gi16553618	Homo sapiens	cDNA FLJ33140 fis, clone UTERU1000160, moderately similar to ZINC FINGER PROTEIN 191.	2173	99
1426	gi10440085	Homo sapiens	cDNA: FLJ23407 fis, clone HEP19601.	1146	100
1426	gi14250146	Homo sapiens	hypothetical protein FLJ23407, clone MGC:14819 IMAGE:4248596, mRNA, complete cds.	1143	99
1427	gi12734104	Homo sapiens	Human DNA sequence from clone RP11-371L19 on chromosome 20. Contains two novel genes, the gene for a novel protein similar to 40S ribosomal protein S10 (RPS10), ESTs, STSs, GSSs and five CpG islands, complete sequence.	2452	100
1427	gi15524116	Homo sapiens	unnamed protein product	2431	98
1427	gi14602488	Homo sapiens	clone MGC:10698 IMAGE:3689286, mRNA, complete cds.	2395	98
1428	AAG67509	Homo sapiens	SMIK Amino acid sequence of a human secreted polypeptide.	4286	100
1428	gi15620867	Homo sapiens	mRNA for KIAA1904 protein, partial cds.	4272	99
1428	gi3191975	Homo sapiens	Human DNA sequence from clone RP1-63G5 on chromosome 22q12.3-13.1 Contains the 3' part of the PSCD4 gene for a human SEC7 homolog B2-1 (cytohesin-2, Arno, ARF exchange factor) LIKE protein, a novel gene and the gene coding for a Leucine rich protein. Contains ESTs, STSs, GSSs and three putative CpG islands, complete sequence.	953	100
1429	AAG02349	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6430.	468	100
1429	gi1770462	Homo sapiens	H.sapiens mRNA for M-phase phosphoprotein, mpp6.	468	100
1429	gi15029628	Homo sapiens	Similar to M-phase phosphoprotein 6, clone MGC:13538 IMAGE:4287267, mRNA, complete cds.	468	100
1430	AAB88377	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0113.	239	100
1430	AAB08904	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 14 SEQ ID NO:61.	239	100
1430	gi142726	Homo sapiens	unnamed protein product	239	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	08				
1431	gi9992893	Homo sapiens	phosphoinositol 3-phosphate binding protein-1 (PEPP1) mRNA, complete cds.	3906	95
1431	AAB42086	Homo sapiens	CURA- Human ORFX ORF1850 polypeptide sequence SEQ ID NO:3700.	427	71
1431	gi4589582	Homo sapiens	mRNA for KIAA0969 protein, complete cds.	256	31
1432	gi9992893	Homo sapiens	phosphoinositol 3-phosphate binding protein-1 (PEPP1) mRNA, complete cds.	4152	99
1432	AAB42086	Homo sapiens	CURA- Human ORFX ORF1850 polypeptide sequence SEQ ID NO:3700.	672	99
1432	AAO12592	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26484.	475	100
1433	AA Y94947	Homo sapiens	GEMY Human secreted protein clone cw1292. 8 protein sequence SEQ ID NO:100.	408	100
1433	AAB65195	Homo sapiens	GETH Human PRO830 (UNQ470) protein sequence SEQ ID NO:175.	215	64
1433	AA Y66672	Homo sapiens	GETH Membrane-bound protein PRO830.	215	64
1434	gi16553818	Homo sapiens	cDNA FLJ25124 fis, clone CBR06414.	1573	100
1434	AAAG02137	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6218.	554	98
1434	AAH75522_aa1	Homo sapiens	SHAN- Human myosin heavy chain 74 encoding cDNA.	275	22
1435	AAF84171_aa1	Homo sapiens	CHUG- Human OATP-B coding sequence.	188	92
1435	AAZ92403_aa1	Homo sapiens	SCHE cDNA encoding human DC-PGT.	188	92
1435	AAC61883_aa1	Homo sapiens	CHIR cDNA encoding a human secreted protein.	188	92
1436	gi11230487	Rattus norvegicus	NTPDase6	501	96
1436	AAB72242	Homo sapiens	HYSE- Mature human CD39 like protein CD39-L2 amino acid sequence.	414	80
1436	AAB72241	Homo sapiens	HYSE- Human CD39 like protein CD39-L2 amino acid sequence.	414	80
1437	gi7243229	Homo sapiens	mRNA for KIAA1424 protein, partial cds.	6604	99
1437	AAB97911	Homo sapiens	SHAN- Human G-protein activating protein 129 SEQ ID NO:2.	6021	99
1437	AAB41660	Homo sapiens	CURA- Human ORFX ORF1424 polypeptide sequence SEQ ID NO:2848.	4377	99
1438	AAB08894	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 4 SEQ ID NO:51.	211	69
1438	gi156262	Buffalopox virus	p8 protein homologue	69	31

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	57				
1438	gi5830678	variola minor virus	A14L protein	68	27
1439	AAG01713	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5794.	670	99
1439	gi12001970	Homo sapiens	clone 015h12 My015 protein mRNA, complete cds.	495	96
1439	gi9963910	Xenopus laevis	Churchill protein	495	71
1440	gi14017831	Homo sapiens	mRNA for KIAA1807 protein, partial cds.	1751	100
1440	gi10438885	Homo sapiens	cDNA: FLJ22479 fis, clone HRC10831.	1524	100
1440	gi14424558	Homo sapiens	KIAA0239 protein, clone IMAGE:4301096, mRNA, partial cds.	157	28
1441	AAB43617	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1062.	1481	93
1441	AAP91913	Homo sapiens	BEHW Anticoagulative PP4X.	1481	93
1441	gi189617	Homo sapiens	Human protein PP4-X mRNA, complete cds.	1481	93
1442	AAE03790	Homo sapiens	HUMA- Human gene 9 encoded secreted protein fragment, SEQ ID NO:60.	391	100
1442	AAE03785	Homo sapiens	HUMA- Human gene 9 encoded secreted protein HMWDW68, SEQ ID NO:55.	391	100
1442	AAAY73425	Homo sapiens	GEMY Human secreted protein clone yj3_1 protein sequence SEQ ID NO:72.	391	100
1443	gi3882329	Homo sapiens	mRNA for KIAA0804 protein, partial cds.	6282	100
1443	AAB94356	Homo sapiens	HELL- Human protein sequence SEQ ID NO:14879.	4093	99
1443	gi10434628	Homo sapiens	cDNA FLJ12883 fis, clone NT2RP2003981, weakly similar to VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN VPS8.	4093	99
1444	gi12539615	Homo sapiens	AKAP-associated sperm protein (ASP) mRNA, complete cds.	1215	99
1444	gi15779077	Homo sapiens	AKAP-associated sperm protein, clone MGC:26950 IMAGE:4820798, mRNA, complete cds.	1212	99
1444	gi11878218	Mus musculus	cAMP-dependent protein kinase regulatory subunit	937	78
1445	AAB43599	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1044.	1283	100
1445	gi14124928	Homo sapiens	clone MGC:3644 IMAGE:2966331, mRNA, complete cds.	1219	100
1445	gi14043853	Homo sapiens	thymidine kinase 1, soluble, clone MGC:14441	1219	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			IMAGE:4303880, mRNA, complete cds.		
1446	AAE00404	Homo sapiens	ZYMO Human phosphodiesterase zcytor13 protein.	2733	100
1446	gi13922371	Homo sapiens	unnamed protein product	2733	100
1446	AAM25548	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1063.	2584	100
1447	gi3766443	Homo sapiens	QDPR gene, exon 1 and joined CDS.	1069	87
1447	gi30819	Homo sapiens	Human mRNA for dihydropteridine reductase (hDHPR).	1069	87
1447	gi181553	Homo sapiens	Human dihydropteridine reductase (hDHPR) mRNA, complete cds.	1069	87
1448	gi13276631	Homo sapiens	mRNA; cDNA DKFZp761F241 (from clone DKFZp761F241); complete cds.	747	100
1448	gi12844872	Mus musculus	putative	650	87
1448	AAV59795	Homo sapiens	META- Human normal ovarian tissue derived protein 72.	554	100
1449	AAB42906	Homo sapiens	CURA- Human ORFX ORF2670 polypeptide sequence SEQ ID NO:5340.	834	100
1449	gi13195151	Homo sapiens	transcription factor T2P (T2P) mRNA, complete cds.	534	54
1449	gi10241461	Homo sapiens	Human DNA sequence from clone RPS-1121G12 on chromosome 20 Contains the 3' end of a gene encoding the hepatocellular carcinoma-associated antigen 58 (HCA58), the SCAND1 gene encoding the domain-containing 1 protein, a novel gene, 2 CpG islands, ESTs, STSs and GSSs, complete sequence.	534	54
1450	AAV12021	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO: 334.	265	97
1450	gi9626218	Beet curly top virus	ORF20.1 > [Beet curly top	63	27
1451	AAG01878	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5959.	348	92
1451	AAB54158	Homo sapiens	HUMA- Human pancreatic cancer antigen protein sequence SEQ ID NO:610.	225	91
1452	AAO08354	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 22246.	1451	89
1452	AAV73384	Homo sapiens	INCY- HTRM clone 2284580 protein sequence.	1451	89
1452	gi13699902	Homo sapiens	mRNA for nucleolar phosphoprotein Nopp34, complete cds.	1451	89
1453	gi142499	Homo sapiens	Similar to hypothetical protein	2387	100

SEQ ID	Hit ID	Species	Description	S score	Percent Identity
	53		FLJ22376, clone MGC:16044 IMAGE:3610443, mRNA, complete cds.		
1453	gi16740559	Homo sapiens	clone MGC:13247 IMAGE:4040497, mRNA, complete cds.	1067	100
1453	gi16551733	Homo sapiens	cDNA FLJ31791 fis, clone NT2RI2008749, weakly similar to SPLICEOSOME ASSOCIATED PROTEIN 49.	1023	53
1454	AAM66321	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26627.	883	50
1454	AAM53933	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26038.	883	50
1454	gi4490388	Felis silvestris	polyprotein	672	44
1455	AAB94815	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15961.	719	100
1455	gi10436205	Homo sapiens	cDNA FLJ13920 fis, clone Y79AA1000420.	719	100
1455	gi12847556	Mus musculus	putative	169	46
1456	gi33044	Homo sapiens	Human mRNA for insulin-like growth factor II (clone P21).	742	97
1456	gi182528	Homo sapiens	Human preproinsulin-like growth factor II (IGF-II) variant mRNA, complete cds.	717	78
1456	AAY70364	Homo sapiens	UYLO- Insulin-like growth factor II.	714	78
1457	AAY99351	Homo sapiens	GETH Human PRO1481 (UNQ750) amino acid sequence SEQ ID NO:41.	1725	100
1457	AAB10259	Homo sapiens	GEMY Human fetal placenta protein fragment BA176 li.	1631	88
1457	AAB10251	Homo sapiens	GEMY Human adult testes protein fragment AJ142 li.	761	97
1458	AAB43607	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1052.	664	88
1458	gi695360	Homo sapiens	nuclear-encoded mitochondrial cytochrome c oxidase Va subunit mRNA, complete cds.	658	87
1458	gi12858580	Mus musculus	putative	544	73
1459	gi7549223	Mus musculus	PALS1	3386	96
1459	AAB94180	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14494.	2590	99
1459	gi10434210	Homo sapiens	cDNA FLJ12615 fis, clone NT2RM4001629, weakly similar to MAGUK P55 SUBFAMILY MEMBER 3.	2590	99
1460	gi12697987	Homo sapiens	mRNA for KIAA1721 protein, partial cds.	3859	99
1460	AAB944	Homo sapiens	HELI- Human protein sequence	3853	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	35		SEQ ID NO:15056.		
1460	gi10434878	Homo sapiens	cDNA FLJ13046 fis, clone NT2RP3001374.	3853	99
1461	gi15214423	Homo sapiens	clone IMAGE:4563921, mRNA, partial cds.	2603	100
1461	gi17901749	Homo sapiens	unnamed protein product	2603	100
1461	gi16740725	Mus musculus	Similar to hexokinase 1	2411	91
1462	AAG75579	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6343.	714	85
1462	AAB43566	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1011.	714	85
1462	gi2398657	Homo sapiens	H.sapiens mRNA translocon-associated protein delta subunit precursor.	714	85
1463	AAG89128	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 248.	2076	100
1463	gi16306850	Homo sapiens	hypothetical protein FLJ22637, clone MGC:2443 IMAGE:2821972, mRNA, complete cds.	2076	100
1463	gi10439104	Homo sapiens	cDNA: FLJ22637 fis, clone HS106677.	2076	100
1464	AAW64262	Homo sapiens	BGHM Human neutrophil elastase.	1326	96
1464	AAP80335	Homo sapiens	TORA) TORAY IND INC (AOKI) Sequence of serine protease (SP) of human myeloid cellorigen and leader peptide.	1326	96
1464	gi386981	Homo sapiens	Human neutrophil elastase gene, exon 5.	1326	96
1465	AAY11385	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID No 207.	220	100
1465	AAB67783	Homo sapiens	INRM Amino acid sequence of a human thyroid NADPH-oxidase.	75	33
1465	AAM24505	Homo sapiens	CORI- Colon tumour related amino acid sequence for C799P.	75	33
1466	gi2895758	Bos taurus	phosphatidic acid-preferring phospholipase A1	4245	91
1466	gi12697955	Homo sapiens	mRNA for KIAA1705 protein, partial cds.	2582	99
1466	gi16554184	Homo sapiens	cDNA FLJ25408 fis, clone TST02965, highly similar to Bos taurus phosphatidic acid-preferring phospholipase A1 mRNA.	2378	100
1467	gi14790025	Homo sapiens	clone MGC:9168 IMAGE:3876839, mRNA, complete cds.	1488	100
1467	gi16768682	Drosophila melanogaster	HL02815p	1155	49
1467	gi10726944	Drosophila melanogaster	CG11306 gene product	1155	49
1468	AAB747	Homo sapiens	HUMA- Human secreted protein	902	91

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	60		sequence encoded by gene 18 SEQ ID NO:69.		
1468	AAB747 59	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 18 SEQ ID NO:68.	902	91
1468	AAB747 50	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 18 SEQ ID NO:59.	902	91
1469	AAB747 60	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 18 SEQ ID NO:69.	1006	96
1469	AAB747 59	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 18 SEQ ID NO:68.	1006	96
1469	AAB747 50	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 18 SEQ ID NO:59.	1006	96
1470	AAW131 08	Homo sapiens	ONYX- Human 14-3-3 beta or HS1 1054.	1225	95
1470	gi279155 2	Homo sapiens	Human DNA sequence from clone RP1-148E22 on chromosome 20q12-13.12 Contains the YWHAB gene encoding tyrosine 3- monooxygenase/tryptophan 5- monooxygenase activation protein, beta polypeptide, a novel gene similar to PABPC1 (poly (A)-binding protein, cytoplasmic 1), 2 CpG islands, ESTs, STSs and GSSs, complete sequence.	1225	95
1470	gi23114	Homo sapiens	H.sapiens mRNA for HS1 protein.	1225	95
1471	AAB948 34	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16000.	2638	100
1471	gi104362 44	Homo sapiens	cDNA FLJ13942 fis, clone Y79AA1000962, weakly similar to MYOSIN HEAVY CHAIN, NON-MUSCLE.	2638	100
1471	gi142905 66	Homo sapiens	hypothetical protein FLJ13942, clone MGC:9884 IMAGE:3867690, mRNA, complete cds.	1501	100
1472	AAB105 50	Homo sapiens	HOFM/ Human aspartate protease psl 4 protein.	1925	100
1472	AAB088 60	Homo sapiens	INCY- Amino acid sequence of a human secretory protein.	1925	100
1472	AAB089 71	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 25 SEQ ID NO:128.	1920	99
1473	AAG892 62	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 382.	231	100
1473	AAV307 21	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	231	100
1473	AAB236 15	Homo sapiens	ALPH- Human secreted protein SEQ ID NO: 30.	222	97
1474	gi702033	Homo sapiens	cDNA FLJ20318 fis, clone	2962	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	6		HEP08704.		
1474	AAM40651	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5582.	1804	47
1474	AAM38865	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2010.	1804	47
1475	gi14495621	Homo sapiens	hypothetical protein FLJ22578, clone MGC:14892 IMAGE:3506508, mRNA, complete cds.	816	100
1475	gi10439014	Homo sapiens	cDNA: FLJ22578 fls, clone HSI02546.	802	100
1475	AAM72825	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 33131.	616	100
1476	gi2224697	Homo sapiens	Human mRNA for KIAA0378 gene, partial cds.	4017	100
1476	gi6681583	Homo sapiens	ELKS mRNA, complete cds.	3463	72
1476	gi13445784	Mus musculus	Rab6-interacting protein 2 isoform A	3423	70
1477	gi15530323	Homo sapiens	clone MGC:4131 IMAGE:2961417, mRNA, complete cds.	3200	99
1477	gi16307502	Mus musculus	Unknown (protein for MGC:11530)	3076	95
1477	gi15277234	Homo sapiens	genomic DNA, chromosome 6p21.3, HLA Class I region, section 12/20.	2227	99
1478	AAY11794	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID No: 394.	375	100
1478	AAB94977	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16558.	68	35
1478	gi10433141	Homo sapiens	cDNA FLJ11800 fls, clone HEMBA1006252.	68	35
1479	AAB08732	Homo sapiens	UYCO Amino acid sequence of a human OLD-35 polypeptide.	3498	98
1479	gi12835817	Mus musculus	putative	2439	89
1479	AAB92684	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11065.	2369	99
1480	gi14424568	Homo sapiens	Mov10 (Moloney leukemia virus 10, mouse) homolog, clone MGC:15000 IMAGE:4109453, mRNA, complete cds.	4983	100
1480	gi12803447	Homo sapiens	Similar to Moloney leukemia virus 10, clone MGC:2948 IMAGE:3138543, mRNA, complete cds.	4983	100
1480	gi10047339	Homo sapiens	mRNA for KIAA1631 protein, partial cds.	4983	100
1481	AAU05584	Homo sapiens	OXFO- Human breast cancer membrane protein 81, BCMP-81.	718	100
1481	AAU25727	Homo sapiens	OXFO- Breast cancer-associated membrane protein (BCMP) 81.	718	100
1481	AAW85738	Homo sapiens	SAGA Polypeptide with transmembrane domain.	718	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1482	gi14017937	Homo sapiens	mRNA for KIAA1870 protein, partial cds.	1496	94
1482	AAB42000	Homo sapiens	CURA- Human ORFX ORF1764 polypeptide sequence SEQ ID NO:3528.	1302	93
1482	AAB93866	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13743.	860	100
1483	gi14017937	Homo sapiens	mRNA for KIAA1870 protein, partial cds.	1608	100
1483	AAB42000	Homo sapiens	CURA- Human ORFX ORF1764 polypeptide sequence SEQ ID NO:3528.	1414	100
1483	AAB93866	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13743.	860	100
1484	gi499184	Felis catus	neuronal protein	617	96
1484	AAB95041	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16804.	507	77
1484	gi7959251	Homo sapiens	mRNA for KIAA1495 protein, partial cds.	507	77
1485	AAB60457	Homo sapiens	INCY- Human cell cycle and proliferation protein CCYPR-5, SEQ ID NO:5.	928	100
1485	gi15080550	Homo sapiens	hypothetical protein FLJ23467, clone MGC:21000 IMAGE:4509736, mRNA, complete cds.	928	100
1485	gi10440166	Homo sapiens	cDNA: FLJ23467 f1s, clone HSI11213.	925	99
1486	AAB93301	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12369.	4341	99
1486	gi14042607	Homo sapiens	cDNA FLJ14812 f1s, clone NTZRP4002081, weakly similar to TRANSCRIPTION INITIATION FACTOR IIA ALPHA AND BETA CHAINS.	4341	99
1486	gi10047179	Homo sapiens	mRNA for KIAA1557 protein, partial cds.	4168	99
1487	AAB94804	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15939.	2050	100
1487	gi15082506	Homo sapiens	hypothetical protein FLJ13910, clone MGC:20406 IMAGE:4636136, mRNA, complete cds.	2050	100
1487	gi10436189	Homo sapiens	cDNA FLJ13910 f1s, clone Y79AA1000131.	2050	100
1488	gi12846013	Mus musculus	putative	1876	97
1488	gi7839559	Homo sapiens	PAD mRNA, complete cds.	1789	98
1488	gi13604169	Homo sapiens	ARG147 mRNA, complete cds.	1575	99
1489	gi12697935	Homo sapiens	mRNA for KIAA1695 protein, partial cds.	2124	100
1489	gi10438624	Homo sapiens	cDNA: FLJ22297 f1s, clone HRC04521.	2124	100
1489	AAB42421	Homo sapiens	CURA- Human ORFX ORF2185 polypeptide sequence	1571	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			SEQ ID NO:4370.		
1490	AAB47562	Homo sapiens	INCY- Protasec PRIS-4.	4321	99
1490	AAM93785	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3804.	4317	99
1490	gi12836332	Mus musculus	putative	4152	95
1491	gi16553816	Homo sapiens	cDNA FLJ25123 fis, clone CBR06154.	1752	93
1491	AAO11834	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 25726.	1347	98
1491	AAM25794	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1309.	919	99
1492	gi14043313	Homo sapiens	clone IMAGE:3609599, mRNA, partial cds.	780	100
1492	AAAY12225	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO: 538.	511	97
1492	AAG00545	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4626.	500	97
1493	AAM93450	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3100.	2693	99
1493	AAAY07754	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 11.	1723	100
1493	AAW79094	Homo sapiens	GEMY Human secreted protein do568 11.	1699	98
1494	AAG64894	Homo sapiens	BIOD- Human phosphoenol pyruvate carboxylase 81.	3851	100
1494	AAB95250	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17415.	1589	100
1494	gi13276665	Homo sapiens	mRNA; cDNA DKFZp761K1524 (from clone DKFZp761K1524); complete cds.	1493	100
1495	AAB43737	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1182.	817	89
1495	AAR59288	Homo sapiens	SHIO Human reg protein.	817	89
1495	gi5764555	Homo sapiens	lithostathine (REG1A) mRNA, complete cds.	817	89
1496	gi12652561	Homo sapiens	Similar to cytochrome b-561, clone MGC:3308 IMAGE:3509626, mRNA, complete cds.	1129	96
1496	gi12804235	Homo sapiens	Similar to cytochrome b-561, clone MGC:2190 IMAGE:3535771, mRNA, complete cds.	1126	95
1496	gi939707	Homo sapiens	Human cytochrome b561 gene, exon 5 and complete cds.	1124	95
1497	gi10437090	Homo sapiens	cDNA: FLJ21077 fis, clone CAS02152.	2182	99
1497	gi10437211	Homo sapiens	cDNA: FLJ21159 fis, clone CAS09969.	1885	100
1497	AAB63960	Homo sapiens	LUDW- Human prostate cancer associated antigen protein sequence SEQ ID NO:1322.	904	96

SEQ ID	HIT ID	Species	Description	S score	Percent identity
1498	AAB40996	Homo sapiens	CURA- Human ORFX ORF760 polypeptide sequence SEQ ID NO:1520.	3391	95
1498	gi10436963	Homo sapiens	cDNA: FLJ20986 fis, clone CAE01156.	3137	99
1498	AAM93525	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3259.	2627	99
1499	AAB99891	Homo sapiens	CHUG- Human RNA helicase gene helicain B protein sequence SEQ ID NO:4.	3818	100
1499	gi16566550	Homo sapiens	DEAD/DEXH helicase DDX31 mRNA, complete cds.	3817	99
1499	gi15215273	Homo sapiens	hypothetical protein FLJ13633, clone MGC:14872 IMAGE:3941452, mRNA, complete cds.	3455	100
1500	gi12005511	Homo sapiens	HT027 mRNA, complete cds.	744	100
1500	gi10436844	Homo sapiens	cDNA: FLJ20886 fis, clone ADKA03257.	739	99
1500	gi11036973	Homo sapiens	HSP22-like protein interacting protein 17 mRNA, complete cds.	459	100
1501	AAB95261	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17444.	906	100
1501	gi10434755	Homo sapiens	cDNA FLJ12967 fis, clone NT2RP2005806.	906	100
1501	AAB94267	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14682.	587	95
1502	gi16533461	Homo sapiens	cDNA FLJ33132 fis, clone UMVEN2000133, weakly similar to RABPHILIN-3A.	2594	99
1502	gi10438690	Homo sapiens	cDNA: FLJ22344 fis, clone HRC06080.	1661	99
1502	AAB93562	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12957.	1289	55
1503	gi16549330	Homo sapiens	cDNA FLJ30165 fis, clone BRACE2000698, weakly similar to ANKYRIN 2.	2280	98
1503	gi12652741	Homo sapiens	clone MGC:3130 IMAGE:3552851, mRNA, complete cds.	2267	78
1503	AAO01850	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15742.	1975	89
1504	AAB66295	Homo sapiens	ICOS- Human tankyrase2 TANK2-SHORT SEQ ID NO: 135.	6088	99
1504	AAB66294	Homo sapiens	ICOS- Human tankyrase2 TANK2-LONG SEQ ID NO: 133.	6088	99
1504	AAB66290	Homo sapiens	ICOS- Human tankyrase2 clone consensus protein SEQ ID NO: 107.	6088	99
1505	gi5689427	Homo sapiens	mRNA for KIAA1045 protein, partial cds.	2087	99
1505	gi13358652	Macaca fascicularis	hypothetical protein	1205	96
1505	ABB1680	Homo sapiens	HUMA- Human nervous system	370	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	6		related polypeptide SEQ ID NO 5463.		
1506	gi5689427	Homo sapiens	mRNA for KIAA1045 protein, partial cds.	2052	94
1506	gi13358652	Macaca fascicularis	hypothetical protein	1205	96
1506	ABB16806	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 5463.	370	100
1507	gi10439066	Homo sapiens	cDNA: FLJ22612 fls, clone HSI04965.	2767	100
1507	gi14529886	Mus musculus	bM145O4.1 (novel protein)	2276	78
1507	gi12855300	Mus musculus	putative	2276	78
1508	gi11493365	Homo sapiens	Human DNA sequence from clone RP5-1009E24 on chromosome 20 Contains the SN gene encoding sialoadhesin, a novel gene similar to KIAA0417, the CENPB gene for centromere protein B, the CDC25B gene for Cell division cycle protein 25B, three novel genes, the 5' end of gene KIAA1271, nine CpG islands, ESTs and GSSs, complete sequence.	6334	99
1508	gi12656130	Homo sapiens	sialoadhesin mRNA, complete cds.	6330	99
1508	gi10440438	Homo sapiens	mRNA for FLJ00055 protein, partial cds.	5046	99
1509	AAAY76539	Homo sapiens	META- Human ovarian tumor EST fragment encoded protein 35.	261	98
1510	AAE04884	Homo sapiens	INCY- Human protense protein-11 (PRTS-11).	424	100
1510	AAB73263	Homo sapiens	UYAL- Human triacylglycerol hydrolase, TGH.	209	51
1510	gi180950	Homo sapiens	Human carboxylesterase mRNA, complete cds.	209	51
1511	AAB94405	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:14983.	4412	100
1511	gi10434798	Homo sapiens	cDNA FLJ12994 fls, clone NT2RP3000207, weakly similar to GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3).	4412	100
1511	gi10437127	Homo sapiens	cDNA: FLJ21104 fls, clone CAS04958.	978	100
1512	AAB62175	Homo sapiens	PLAC Human p110FYVE protein.	4028	99
1512	AAF57403_aa1	Homo sapiens	PLAC Human p110FYVE protein encoding DNA.	4027	99
1512	gi11344951	Homo sapiens	FYVE-finger-containing Rab5 effector protein Rabenosyn-5 mRNA, complete cds.	4027	99
1513	gi795926	Homo sapiens	mRNA for KIAA1500 protein,	4061	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	1		partial cds.		
1513	gi9837427	Lytechinus variegatus	embryonic blastocoelar extracellular matrix protein precursor	1085	34
1513	AAG73354	Homo sapiens	HUMA- Human gene 9-encoded secreted protein HETAM53, SEQ ID NO:125.	517	100
1514	gi16550108	Homo sapiens	cDNA FLJ30829 fls, clone FEBRA2001790, highly similar to Xenopus laevis RRM-containing protein SEB-4 mRNA.	914	100
1514	gi13624461	Homo sapiens	Human DNA sequence from clone RP1-259A10 on chromosome 6p22.1-23 Contains the gene for an ssDNA binding protein (SEB4D), ESTs, STSs, GSSs and a CpG island, complete sequence.	914	100
1514	gi8895698	Xenopus laevis	RRM-containing protein SEB-4	790	88
1515	gi16551580	Homo sapiens	cDNA FLJ31673 fls, clone NT2RI2005061.	3158	99
1515	AAB94929	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16399.	2275	100
1515	gi10432847	Homo sapiens	cDNA FLJ11565 fls, clone HEMBA1003229.	2275	100
1516	gi3947688	Homo sapiens	mRNA for Sec24 protein (Sec24A isoform), partial.	5355	98
1516	AAM79111	Homo sapiens	HYSE- Human protein SEQ ID NO 1773.	3090	55
1516	gi3947690	Homo sapiens	mRNA for Sec24 protein (Sec24B isoform).	3090	55
1517	AAAY12049	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO: 362.	253	92
1517	gi4155063	Helicobacter pylori J99	putative	75	29
1517	gi5738522	Schizosaccharomyces pombe	putative pre-mrna splicing factor atp-dependent rna helicase	69	33
1518	AAB20349	Homo sapiens	UYRQ Human vomeronasal-like receptor hVLR1 (long form).	1859	99
1518	AAG64295	Homo sapiens	HELI- Human GTP-binding protein-coupled receptor GPRv31.	1859	99
1518	gi9988585	Homo sapiens	putative pheromone receptor V1RL1 long form (V1RL1) mRNA, complete cds.	1859	99
1519	gi1424909	Homo sapiens	clone IMAGE:3506174, mRNA, partial cds.	2759	90
1519	gi1424907	Homo sapiens	clone IMAGE:3506145, mRNA, partial cds.	2759	90
1519	AAAY99355	Homo sapiens	GETH Human PRO1295 (UNQ664) amino acid sequence SEQ ID NO:54.	1265	100
1520	gi16041686	Homo sapiens	hypothetical protein FLJ22393, clone MGC:16798 IMAGE:3916157, mRNA,	1470	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			complete cds.		
1520	gi10438763	Homo sapiens	cDNA: FLJ22393 fls, clone HRC07880.	1463	99
1520	gi12853419	Mus musculus	putative	1446	98
1521	AAR35072	Homo sapiens	UYPR- Human t-complex associated testes expressed protein 1.	2576	97
1521	gi201910	Mus musculus	Tcte-1 peptide	1883	74
1521	gi7300285	Drosophila melanogaster	CG14325 gene product	348	27
1522	gi12851762	Mus musculus	putative	689	88
1522	AAG02298	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6379.	382	100
1522	gi16769034	Drosophila melanogaster	LD15209p	294	38
1523	AAQ90526.na1	Homo sapiens	OKLA- Human SIII 15 kDa subunit cDNA.	426	100
1523	AAW13850	Homo sapiens	OKLA- Human RNA polymerase transcription factor elongin 15 kDa subunit.	426	100
1523	AAR75087	Homo sapiens	OKLA- Human SIII 15 kDa subunit.	426	100
1524	gi12855672	Mus musculus	putative	2165	85
1524	AAU17429	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 994.	987	98
1524	AAG04081	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8162.	507	99
1525	AAW13085	Homo sapiens	SAGA Human E2 ubiquinone binding enzyme.	667	88
1525	gi13097195	Homo sapiens	ubiquitin-conjugating enzyme E2N (homologous to yeast UBC13), clone MGC:5063 IMAGE:2906313, mRNA, complete cds.	667	88
1525	gi12653255	Homo sapiens	ubiquitin-conjugating enzyme E2N (homologous to yeast UBC13), clone MGC:8489 IMAGE:2822013, mRNA, complete cds.	667	88
1526	AAV87271	Homo sapiens	INCY- Human signal peptide containing protein HSPP-48 SEQ ID NO:48.	471	86
1526	gi17128086	Corynebacterium glutamicum	cdsA	70	27
1526	gi12544226	Corynebacterium glutamicum	RXA01894	70	27
1527	gi13358942	Macaca fascicularis	hypothetical protein	2660	97
1527	AAB87587	Homo sapiens	GETH Human PRO1693.	2647	100
1527	AAU12439	Homo sapiens	GETH Human PRO1693 polypeptide sequence.	2647	100
1528	AAB425	Homo sapiens	CURA- Human ORFX	2137	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	73		ORF2337 polypeptide sequence SEQ ID NO:4674.		
1528	gi12845823	Mus musculus	putative	1792	95
1528	gi4495063	Homo sapiens	Human DNA sequence from clone 108K11 on chromosome 6p21. Contains SRP20 (SR protein family member), Ndr protein kinase gene similar to yeast suppressor protein SRP40, EST and GSS, complete sequence.	1468	99
1529	gi15823636	Homo sapiens	ALS2 mRNA, complete cds, long form.	8660	99
1529	gi16076812	Homo sapiens	alsin mRNA, complete cds.	8646	99
1529	gi15823640	Mus musculus	Als2	8005	91
1530	AA664171	Homo sapiens	TAKE Human profilin IIL	750	100
1530	gi12804213	Homo sapiens	profilin 2, clone MGC:1684 IMAGE:3533907, mRNA, complete cds.	750	100
1530	gi10952520	Homo sapiens	profilin 1a (PFN2) mRNA, complete cds, alternatively spliced.	750	100
1531	AA664171	Homo sapiens	TAKE Human profilin IIL	636	87
1531	gi12804213	Homo sapiens	profilin 2, clone MGC:1684 IMAGE:3533907, mRNA, complete cds.	636	87
1531	gi10952520	Homo sapiens	profilin 1a (PFN2) mRNA, complete cds, alternatively spliced.	636	87
1532	AAB94952	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:16482.	1830	99
1532	gi17907791	Homo sapiens	TAIP-2 mRNA for TGF-beta induced apoptosis protein 2, complete cds.	1830	99
1532	gi10433016	Homo sapiens	cDNA FLJ11703 fis, clone HEMBA1005075.	1830	99
1533	gi14133223	Homo sapiens	mRNA for KIAA0876 protein, partial cds.	4559	100
1533	gi6910563	Homo sapiens	chromosome 19, BC335474 (CIT-HSPC_482H14), complete sequence.	4370	99
1533	gi13938056	Mus musculus	Similar to KIAA0677 gene product	3313	73
1534	gi7243191	Homo sapiens	mRNA for KIAA1405 protein, partial cds.	3986	99
1534	gi12313647	Mus musculus	MmKIF17	3319	77
1534	gi4102184	Homo sapiens	KIF3-related motor protein (KIF3X) mRNA, partial cds.	1084	90
1535	gi12053311	Homo sapiens	mRNA; cDNA DKFZp434K229 (from clone DKFZp434K229); complete cds.	1600	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1535	gi10438986	Homo sapiens	cDNA: FLJ22557 fis, clone HSI01483.	1405	100
1535	gi15214537	Mus musculus	RIKEN cDNA 2410042D21 gene	1369	86
1536	gi12274933	Homo sapiens	mRNA for alanine:glyoxylate aminotransferase 2 homolog 1, splice form 1 (AGXT2L1 gene).	2018	100
1536	gi12836724	Mus musculus	putative	1689	83
1536	gi15859690	Homo sapiens	unnamed protein product	1189	66
1537	gi14017847	Homo sapiens	mRNA for KIAA1815 protein, partial cds.	2117	100
1537	gi12314159	Homo sapiens	Human DNA sequence from clone RP11-207C16 on chromosome 9p23-24.3. Contains the 3' end of the gene for a novel protein similar to C. elegans R06F6.8 (Sw:Q09417) (contains KIAA1432), the 3' end of the gene for a novel protein similar to predicted yeast, plant and worm proteins, ESTs, STSs and GSSs, complete sequence.	2117	100
1537	gi10439948	Homo sapiens	cDNA: FLJ23309 fis, clone HEP11618.	1725	99
1538	gi10437187	Homo sapiens	cDNA: FLJ21144 fis, clone CAS07955.	1919	99
1538	AAB95360	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17664.	1912	99
1538	gi12851810	Mus musculus	putative	1672	84
1539	gi3327036	Homo sapiens	mRNA for KIAA0611 protein, partial cds.	4702	100
1539	gi13905302	Mus musculus	Similar to ATPase, class II, type 9A	3961	98
1539	gi6434968	Mus musculus	putative E1-E2 ATPase	3942	98
1540	AAB93976	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14051.	2580	100
1540	gi10433868	Homo sapiens	cDNA FLJ12401 fis, clone MAMMA1002796.	2580	100
1540	AAM69800	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 30106.	1108	100
1541	AAB94901	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16290.	1317	99
1541	gi10432716	Homo sapiens	cDNA FLJ11457 fis, clone HEMBA1001522.	1317	99
1541	gi12853191	Mus musculus	putative	887	62
1542	AAB73507	Homo sapiens	INCY- Human transferase HTFS-14, SEQ ID NO:14.	1698	99
1542	gi16552496	Homo sapiens	cDNA FLJ32390 fis, clone SKMUS1000177, weakly similar to PROTEIN-L-ISOASPARTATE O-	1698	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			METHYLTRANSFERASE (EC 2.1.1.77).		
1542	gi11323192	Homo sapiens	Human DNA sequence from clone RP5-1022E24 on chromosome 20 Contains the 3' end of the OPRL1 gene encoding Opiate receptor-like 1 protein, the GPR8 gene encoding a G protein-coupled receptor, the KIAA0835 gene encoding a protein similar to the myelin transcription factor 1 (MYT1), a novel gene, 7 CpG islands, ESTs, STSs and GSSs, complete sequence.	1156	69
1543	AAB94644	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15536.	1941	100
1543	gi14625875	Homo sapiens	mRNA for putative autophagy-related cysteine endopeptidase (AUTL1 gene).	1941	100
1543	gi14042698	Homo sapiens	cDNA FLJ14867 fis, clone PLACE1002319.	1941	100
1544	gi9967204	Macaca fascicularis	hypothetical protein	2621	100
1544	AAB88351	Homo sapiens	HELI- Human membrane or secretory protein clone PSEC0076.	2616	99
1544	gi14272556	Homo sapiens	unnamed protein product	2616	99
1545	gi12005429	Homo sapiens	homeobox-containing transcription factor HOXD1 (HOXD1) mRNA, complete cds.	1726	100
1545	gi11095618	Homo sapiens	HOX D1 protein (HOXD1) gene, complete cds.	1726	100
1545	gi15680245	Homo sapiens	homeo box D1, clone MGC:23144 IMAGE:4869019, mRNA, complete cds.	1718	99
1546	AAB95366	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17677.	2511	100
1546	gi14042538	Homo sapiens	cDNA FLJ14773 fis, clone NT2RP3004566, weakly similar to ZINC FINGER PROTEIN 84.	2511	100
1546	gi10438630	Homo sapiens	cDNA: FLJ22301 fis, clone HRC04777.	2511	100
1547	gi15559628	Homo sapiens	mitochondrial ribosomal protein S5, clone MGC:20735 IMAGE:4561399, mRNA, complete cds.	2262	99
1547	gi13620881	Homo sapiens	MRPS5 mRNA for mitochondrial ribosomal protein S5, complete cds.	2262	99
1547	gi13620883	Mus musculus	mitochondrial ribosomal protein S5	1821	79
1548	gi12697911	Homo sapiens	mRNA for KIAA1683 protein, partial cds.	1772	72
1548	gi12053239	Homo sapiens	mRNA; cDNA DKFZp434O194 (from clone DKFZp434O194); complete cds.	1690	49

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1548	AAU17484	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 1049.	1659	100
1549	gi10438166	Homo sapiens	cDNA: FLJ21945 fts, clone HEP04702.	3738	99
1549	gi10441956	Homo sapiens	clone PP384 unknown mRNA.	1196	99
1549	AAU22252	Homo sapiens	HUMA- Human cardiovascular system antigen polypeptide SEQ ID No 1026.	480	98
1550	gi12852481	Mus musculus	putative	1049	56
1550	AAB18966	Homo sapiens	INCY- Amino acid sequence of a human transmembrane protein.	922	89
1550	AAB42781	Homo sapiens	CURA- Human ORFX ORF2545 polypeptide sequence SEQ ID NO:5090.	513	100
1551	gi6330401	Homo sapiens	mRNA for KIAA1199 protein, partial cds.	2521	97
1551	AAAY25793	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 12.	1919	96
1551	gi8518188	Homo sapiens	mRNA full length insert cDNA clone EUROIMAGE 2155353.	1352	95
1552	AAB94417	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15016.	1027	99
1552	gi10434835	Homo sapiens	cDNA FLJ13018 fts, clone NT2RP300685.	1027	99
1552	gi12837567	Mus musculus	putative	972	92
1553	gi15929411	Homo sapiens	clone IMAGE:4040789, mRNA, partial cds.	1369	100
1553	AAG03490	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7571.	450	96
1553	gi12845828	Mus musculus	putative	376	68
1554	gi9988958	Homo sapiens	F-box protein FBX30 mRNA, complete cds.	1211	100
1554	gi14043744	Homo sapiens	Similar to F-box only protein 6, clone MGC:14140 IMAGE:4054414, mRNA, complete cds.	1211	100
1554	AAB42989	Homo sapiens	CURA- Human ORFX ORF2753 polypeptide sequence SEQ ID NO:5506.	675	99
1555	gi938026	Homo sapiens	Human mRNA for RanBP1 (Ran-binding protein 1), complete cds.	916	100
1555	AAB56619	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1197.	904	99
1555	gi620083	Homo sapiens	H.sapiens mRNA for RanBP1.	904	99
1556	gi13383265	Homo sapiens	mRNA for actin related protein, complete cds.	1962	100
1556	gi16550055	Homo sapiens	cDNA FLJ30784 fts, clone FEBRA2000881, moderately similar to ACTIN 6.	1950	99
1556	gi139383	Homo sapiens	clone MGC:15664	1949	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	19		IMAGE:3349184, mRNA, complete cds.		
1557	gi623545	Oryctolagus cuniculus	sarcoplasmic reticulum glycoprotein	2366	96
1557	gi164861	Oryctolagus cuniculus	sarcolumenin precursor	2307	97
1557	gi496325	Gallus gallus	53 kDa glycoprotein	2160	87
1558	AAB94641	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15526.	1379	98
1558	AAG64403	Homo sapiens	SHAN- Human paneth cell enhanced expression-like protein.	1379	98
1558	AAM94028	Homo sapiens	HELI- Human stomach cancer expressed polypeptide SEQ ID NO 126.	1379	98
1559	gi8745315	Homo sapiens	putative GTP-binding protein (GTPBP2) mRNA, partial cds.	2742	99
1559	gi13561007	Homo sapiens	Human DNA sequence from clone RP11-22124 on chromosome 6 Contains the 3' part of the POLH gene for DNA directed polymerase α and the GTPBP2 gene for GTP binding protein 2, ESTs, STSs, GSSs and a CpG Island, complete sequence.	2742	99
1559	gi8745317	Mus musculus	putative GTP-binding protein	2731	99
1560	AAG02653	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6734.	425	100
1560	gi15080459	Homo sapiens	clone MGC-9017 IMAGE:3860059, mRNA, complete cds.	425	100
1560	gi2337920	Homo sapiens	Human syntaxin 7 mRNA, complete cds.	421	98
1561	AAB94468	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15128.	4375	99
1561	gi10434944	Homo sapiens	cDNA FLJ13089 fls, clone NT2RP3002108, weakly similar to DEC1 PROTEIN.	4375	99
1561	gi7300581	Drosophila melanogaster	CG4845 gene product	1083	31
1562	AAU12177	Homo sapiens	GETH Human PRO305 polypeptide sequence.	226	100
1562	AAV64734	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:895.	226	100
1562	AAV81487	Homo sapiens	FUJY Human cathepsin L2.	226	100
1563	gi12052726	Homo sapiens	mRNA; cDNA DKFZp761N0411 (from clone DKFZp761N0411); complete cds.	2381	99
1563	gi14250920	Homo sapiens	mRNA for SMC6 protein.	2374	99
1563	gi14250922	Mus musculus	SMC6 protein	2163	88
1564	gi149705	Homo sapiens	mRNA for WDR9 protein	524	98

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	62		(WDR9 gene), form A.		
1564	gi14970591	Mus musculus	WDR9 protein, form A	363	79
1564	AAB34487	Homo sapiens	HUMA- Human secreted protein BLAST search protein SEQ ID NO: 105.	169	52
1565	gi6808329	Homo sapiens	mRNA; cDNA DKFZp434K0410 (from clone DKFZp434K0410); partial cds.	787	100
1565	AAB93188	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12140.	528	51
1565	AAB92702	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11102.	528	51
1566	gi12803841	Homo sapiens	Similar to retinoic acid induced 12, clone MGC:3373 IMAGE:3629369, mRNA, complete cds.	1576	99
1566	AAB26793	Homo sapiens	UYFU- Human melanoma growth related factor-1 amino acid sequence.	1199	95
1566	gi13277702	Mus musculus	retinoic acid induced 12	1182	75
1567	AAB53405	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:945.	1288	93
1567	gi29587	Homo sapiens	Human mRNA for carbonic anhydrase II (EC 4.2.1.1).	1288	93
1567	gi179795	Homo sapiens	Human carbonic anhydrase II mRNA, complete cds.	1288	93
1568	gi182146	Homo sapiens	eosinophil peroxidase (EPP) gene, exon 12 and complete cds.	3757	100
1568	gi31183	Homo sapiens	Human mRNA for eosinophil peroxidase.	3549	97
1568	gi1777378	Mus musculus	eosinophil peroxidase	3376	89
1569	AAB94183	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14500.	2405	100
1569	gi10434215	Homo sapiens	cDNA FLJ12618 fis, clone NT2RM4001666, weakly similar to HYPOTHETICAL 48.6 KD PROTEIN IN ALPA-GABP INTERGENIC REGION.	2405	100
1569	gi13543955	Homo sapiens	Similar to hypothetical protein FLJ12618, clone MGC:12994 IMAGE:3504996, mRNA, complete cds.	2082	94
1570	AAB93904	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13862.	2499	99
1570	gi14286214	Homo sapiens	hypothetical protein FLJ12150, clone MGC:15043 IMAGE:3634992, mRNA, complete cds.	2499	99
1570	gi1043359	Homo sapiens	cDNA FLJ12150 fis, clone MAMMA1000422.	2499	99
1571	gi17684605	Mus musculus	smoothelin B	640	54
1571	gi1768460	Mus musculus	smoothelin A	640	54

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	4				
1571	gi7547258	Mus musculus	smoothelin small isoform S1	640	54
1572	AAB95033	Homo sapiens	HELL- Human protein sequence SEQ ID NO:16784	1174	100
1572	gi10433442	Homo sapiens	cDNA FLJ12056 fis, clone HEMBB1002050.	1174	100
1572	gi530876	Chlamydomonas reinhardtii	amino acid feature: Rod protein domain, aa 266 .. 468; amino acid feature: globular protein domain, aa 32 .. 265	142	26
1573	gi12857247	Mus musculus	putative	2111	92
1573	gi10727909	Drosophila melanogaster	CG6169 gene product	688	48
1573	gi2388911	Schizosaccharomyces pombe	hypothetical PSU1-like protein	585	47
1574	AAB41954	Homo sapiens	CURA- Human ORFX ORF1718 polypeptide sequence SEQ ID NO:3436	1765	97
1574	AAB42773	Homo sapiens	CURA- Human ORFX ORF2537 polypeptide sequence SEQ ID NO:5074	1134	93
1574	gi17512254	Homo sapiens	hypothetical protein FLJ21156, clone MGC:29459 IMAGE:5020837, mRNA, complete cds.	1089	100
1575	AAY40090	Homo sapiens	HUMA- Peptide sequence derived from a human secreted protein.	918	98
1575	gi12854639	Mus musculus	putative	443	69
1575	gi17066107	Homo sapiens	partial TTN gene for titin.	86	25
1576	gi10438473	Homo sapiens	cDNA: FLJ22184 fis, clone HRC00983.	3291	99
1576	gi10241712	Homo sapiens	mRNA; cDNA DKFZp761K0816 (from clone DKFZp761K0816).	1238	99
1576	gi600118	Zen mays	extensin-like protein	666	33
1577	gi16549261	Homo sapiens	cDNA FLJ30107 fis, clone BNGH41000198, weakly similar to TETRACYCLINE RESISTANCE PROTEIN, CLASS E.	1453	100
1577	gi14715055	Homo sapiens	Similar to RIKEN cDNA 1110002C08 gene, clone MGC:9564 IMAGE:3872267, mRNA, complete cds.	1453	100
1577	gi12833845	Mus musculus	putative	1339	90
1578	gi10437669	Homo sapiens	cDNA: FLJ21551 fis, clone COL06266.	1925	99
1578	AAE01791	Homo sapiens	HUMA- Human gene 22 encoded secreted protein HOHDF66, SEQ ID NO:112.	1840	99
1578	AAB417	Homo sapiens	CURA- Human ORFX	1473	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	11		ORF1475 polypeptide sequence SEQ ID NO:2950.		
1579	gi14327915	Homo sapiens	clone MGC:11186 IMAGE:3844322, mRNA, complete cds.	3446	100
1579	gi14043103	Homo sapiens	clone MGC:15388 IMAGE:3350378, mRNA, complete cds.	3446	100
1579	AAB94722	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15741.	3439	99
1580	gi13446190	Homo sapiens	Human DNA sequence from clone RP4-717M23 on chromosome 20 Contains the gene encoding a CRP2 binding protein (CRP2BP), a pseudogene, ESTs, STSs, GSSs and CpG islands, complete sequence.	3955	97
1580	gi14043103	Homo sapiens	clone MGC:15388 IMAGE:3350378, mRNA, complete cds.	3952	97
1580	gi14327915	Homo sapiens	clone MGC:11186 IMAGE:3844322, mRNA, complete cds.	3281	96
1581	gi16551610	Homo sapiens	cDNA FLJ31697 fis, clone NT2RI2005851, weakly similar to PLECTIN.	1911	99
1581	AAM66435	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26741.	588	100
1581	AAM54044	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26149.	588	100
1582	AAU20443	Homo sapiens	HUMA- Human secreted protein, Seq ID No 435.	940	94
1582	AAM93713	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3654.	756	100
1582	gi17149039	Homo sapiens	MTOL-like protein gene, complete cds; nuclear gene for mitochondrial product.	756	100
1583	ABB12220	Homo sapiens	HYSE- Human peroxisomal Ca- dependent solute carrier homologue, SEQ:2590.	344	100
1583	gi12853685	Mus musculus	putative	168	52
1583	AAM80061	Homo sapiens	HYSE- Human protein SEQ ID NO 3707.	165	55
1584	gi14349125	Homo sapiens	mRNA for alpha2- glucosyltransferase (ALG10 gene).	716	93
1584	gi3513451	Rattus norvegicus	potassium channel regulator 1	681	90
1584	AAB25715	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 3 SEQ ID NO:104.	617	93
1585	gi12697939	Homo sapiens	mRNA for KIAA1697 protein, partial cds.	1904	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1585	gi10440249	Homo sapiens	cDNA: FLJ23529 fls, clone LNG06042.	1897	99
1585	gi7293415	Drosophila melanogaster	Dhc16F gene product	786	44
1586	AAZ35834_aa1	Homo sapiens	INCY- Human vesicle trafficking protein 2 encoding cDNA.	804	84
1586	AAB93664	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13188.	804	84
1586	AAAY49959	Homo sapiens	INCY- Human vesicle trafficking protein 2.	804	84
1587	AAAY76561	Homo sapiens	META- Human ovarian tumor EST fragment encoded protein 57.	623	92
1587	gi13623585	Homo sapiens	Similar to RIKEN cDNA 1500034E06 gene, clone MGC:14151 IMAGE:3690202, mRNA, complete cds.	623	92
1587	gi12858676	Mus musculus	putative	595	88
1588	AAU20647	Homo sapiens	HUMA- Human secreted protein, Seq ID No 639.	927	99
1588	AAU20523	Homo sapiens	HUMA- Human secreted protein, Seq ID No 515.	927	99
1588	AAB95432	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17854.	927	99
1589	AAB59191	Homo sapiens	UYCO Human NADE.	293	57
1589	gi8452894	Homo sapiens	p75NTR-associated cell death executor (NADE) mRNA, complete cds.	293	57
1589	gi189379	Homo sapiens	Human unknown protein from clone pHGR74 mRNA, complete cds.	293	57
1590	AAAG01716	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5797.	894	99
1590	AAAY82473	Homo sapiens	LLTL- Human APG12 protein sequence.	724	100
1590	gi4115731	Homo sapiens	mRNA for Apg12, complete cds.	724	100
1591	AAAW87701	Homo sapiens	INCY- A human membrane fusion protein designated SYTX1.	1357	99
1591	gi4200241	Homo sapiens	H.sapiens gene from PAC 42616, similar to syntaxin 7.	1325	100
1591	gi14715019	Mus musculus	Unknown (protein for MGC:6471)	1280	93
1592	AAA54089_aa1	Homo sapiens	GETH PRO211 cDNA.	1944	87
1592	AAB53075	Homo sapiens	GETH Human angiogenesis-associated protein PRO211, SEQ ID NO:57.	1944	87
1592	AAB61231	Homo sapiens	MILL- Human TANGO 331 protein.	1944	87
1593	gi6681592	Homo sapiens	HSJ2 mRNA for DnaJ homolog, complete cds.	1567	93
1593	gi128032	Homo sapiens	MRJ gene for a member of the	1567	93

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	63		DNAJ protein family, clone MGC:1152 IMAGE:3346070, mRNA, complete cds.		
1593	AAW94066	Homo sapiens	INCY- Human DnaJ-like protein, HSPJ2.	1516	98
1594	gi1230858	Homo sapiens	mRNA for HMG-box transcription factor TCF-3, complete cds.	3096	100
1594	gi123638	Mus musculus	TCF-3 protein	2934	95
1594	gi14279982	Xenopus laevis	T-cell factor XTCTF-3	2195	77
1595	gi16307074	Homo sapiens	hypothetical protein FLJ22724, clone MGC:16791 IMAGE:3900548, mRNA, complete cds.	932	100
1595	gi10439225	Homo sapiens	cDNA: FLJ22724 fis, clone HSI14868.	932	100
1595	gi12854396	Mus musculus	putative	618	68
1596	gi10047229	Homo sapiens	mRNA for KIAA1577 protein, partial cds.	3874	100
1596	gi7959283	Homo sapiens	mRNA for KIAA1511 protein, partial cds.	3002	74
1596	gi10440418	Homo sapiens	mRNA for FLJ00044 protein, partial cds.	2010	63
1597	AAAY1317	Homo sapiens	GEST Human secreted protein encoded by 5' EST SEQ ID NO: 131.	268	100
1597	gi13816407	Sulfolobus solfataricus	Dehydrogenase, putative	66	43
1597	gi15140082	Sinorhizobium meliloti	HYPOTHETICAL PROTEIN	66	37
1598	gi16943720	Homo sapiens	mRNA for SOX7 protein.	2106	100
1598	gi16550314	Homo sapiens	cDNA FLJ30994 fis, clone HLUNG1000076, highly similar to Mus musculus mRNA for mSox7.	2106	100
1598	gi13279164	Homo sapiens	Similar to SRY-box containing gene 7, clone MGC:10895 IMAGE:3622936, mRNA, complete cds.	2106	100
1599	gi17646146	Homo sapiens	B lymphocyte activation-related protein mRNA, complete cds.	1111	97
1599	gi12006223	Homo sapiens	NPD017 mRNA, complete cds.	1111	97
1599	gi15126745	Homo sapiens	hypothetical protein FLJ21174, clone MGC:5372 IMAGE:3445403, mRNA, complete cds.	1105	97
1600	gi14328879	Homo sapiens	CUB domain containing protein 1 (CDCP1) mRNA, complete cds.	4394	100
1600	gi10439515	Homo sapiens	cDNA: FLJ22969 fis, clone KAT10759.	4385	99
1600	AAAY914	Homo sapiens	HUMA- Human secreted protein	3633	98

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	56		sequence encoded by gene 6 SEQ ID NO:129.		
1601	gi12052846	Homo sapiens	mRNA; cDNA DKFZp564K2464 (from clone DKFZp564K2464); complete cds.	2152	100
1601	gi12652723	Homo sapiens	clone MGC:3295 IMAGE:3508204, mRNA, complete cds.	2035	100
1601	AAW79088	Homo sapiens	GEMY Human secreted protein bil29_2.	1115	100
1602	gi5081610	Mus musculus	huntington yeast partner C	4295	94
1602	gi15636898	Gallus gallus	formin binding protein 11- related protein	2507	55
1602	gi5081608	Mus musculus	formin binding protein 11	2505	54
1603	gi5081610	Mus musculus	huntington yeast partner C	4046	90
1603	gi6808038	Homo sapiens	mRNA; cDNA DKFZp434H2121 (from clone DKFZp434H2121); partial cds.	2341	100
1603	gi15636898	Gallus gallus	formin binding protein 11- related protein	2339	52
1604	AAB95053	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16855.	680	100
1604	gi10433525	Homo sapiens	cDNA FLJ12122 fis, clone MAMMA1000129.	680	100
1604	AAE06096	Homo sapiens	HUMA- Human gene 56 encoded secreted protein HRABA80, SEQ ID NO:158.	151	48
1605	AAB94309	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14777.	2318	99
1605	gi10434501	Homo sapiens	cDNA FLJ12800 fis, clone NT2RP2002079, weakly similar to HISTONE H1, GONADAL.	2318	99
1605	gi1739125	Mus musculus	Similar to hypothetical protein FLJ12800	1515	70
1606	AAG64502	Homo sapiens	UYFU- Human EPIP2 (endometrial progesterone- induced protein).	1914	100
1606	gi17390289	Homo sapiens	Similar to phosphoserine aminotransferase, clone MGC:9290 IMAGE:3883843, mRNA, complete cds.	1914	100
1606	gi13436074	Homo sapiens	Similar to phosphoserine aminotransferase, clone MGC:10519 IMAGE:3938160, mRNA, complete cds.	1914	100
1607	gi5326802	Homo sapiens	phosphoserine aminotransferase (PSA) mRNA, complete cds.	1673	100
1607	AAG64502	Homo sapiens	UYFU- Human EPIP2 (endometrial progesterone- induced protein).	1616	87
1607	gi17390289	Homo sapiens	Similar to phosphoserine aminotransferase, clone MGC:9290 IMAGE:3883843,	1616	87

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			mRNA, complete cds.		
1608	gi13447761	Homo sapiens	cystatin and DUF19 domain-containing protein 1 (CSDUF1) mRNA, complete cds.	736	100
1608	AAG74513	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5277.	570	100
1608	AAB93798	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13564.	281	47
1609	gi10440230	Homo sapiens	cDNA: FLJ23514 fis, clone LNG04628.	2005	100
1609	gi12852973	Mus musculus	putative	1509	69
1609	gi12407749	Arabidopsis thaliana	initiation factor 3a	138	22
1610	AAV64994	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:1155.	372	100
1610	AAM00852	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 215.	69	39
1610	gi332565	Orf virus	ORF2	68	42
1611	gi10439705	Homo sapiens	cDNA: FLJ23121 fis, clone LNG07996.	3137	100
1611	AAB94996	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16625.	1618	99
1611	gi10433257	Homo sapiens	cDNA FLJ11889 fis, clone HEMBA1007251, weakly similar to Homo sapiens F-box protein FBX29 (FBX29) mRNA.	1618	99
1612	AAB95234	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17375.	3584	100
1612	gi10434674	Homo sapiens	cDNA FLJ12911 fis, clone NT2RP2004425, highly similar to Mus musculus axotrophin mRNA.	3584	100
1612	gi5052031	Mus musculus	axotrophin	2983	85
1613	gi9650954	Mus musculus	beta-1,6-N-acetylglucosaminyltransferase B	1254	73
1613	AAV16000_aa1	Homo sapiens	LJOL- Human beta-1,6-N-acetylglucosaminyltransferase (IGNT) encoding cDNA.	1044	65
1613	AAQ89201_aa1	Homo sapiens	LJOL- l-branching enzyme cDNA.	1044	65
1614	gi10047311	Homo sapiens	mRNA for KIAA1617 protein, partial cds.	4792	100
1614	gi6635353	Homo sapiens	RU1 (RU1) mRNA, complete cds.	2467	55
1614	gi15779095	Homo sapiens	Similar to RU1, clone MGC:3342 IMAGE:3029598, mRNA, complete cds.	2467	55
1615	gi4680693	Homo sapiens	CGI-27 protein mRNA, complete cds.	1414	92
1615	gi17511762	Homo sapiens	CGI-27 protein, clone MGC:31852 IMAGE:4851517, mRNA, complete cds.	1414	92
1615	gi170463	Homo sapiens	C21orf19-like protein mRNA,	1414	92

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	03		complete cds.		
1616	gi10047269	Homo sapiens	mRNA for KIAA1597 protein, partial cds.	4638	99
1616	gi13647069	Mus musculus	synaptotagmin-like protein 2-a delta 2S-III	3717	81
1616	gi13647009	Mus musculus	synaptotagmin-like protein 2-a	3666	77
1617	AAM93772	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3778.	1945	98
1617	ABB11731	Homo sapiens	HYSE- Human granuphilin-a homologue, SEQ ID NO:2101.	1945	98
1617	gi15930218	Homo sapiens	synaptotagmin-like 2, clone MGC:9588 IMAGE:3887570, mRNA, complete cds.	1945	98
1618	gi7768739	Homo sapiens	genomic DNA, chromosome 21q, section 89/105.	3747	100
1618	gi12857381	Mus musculus	putative	1233	78
1618	gi4884386	Homo sapiens	mRNA; cDNA DKFZp586F0422 (from clone DKFZp586F0422); partial cds.	870	80
1619	AAR22546	Homo sapiens	NEUR- Truncated Dopamine D1 receptor encoded by pseudogene clone GL-39.	1157	93
1619	AAR21082	Homo sapiens	NEUR- Dopamine D1 receptor encoded by clone GL-30.	1028	85
1619	gi32049	Homo sapiens	Human HD5DR gene for D5 dopamine receptor.	1028	85
1620	AAE10329	Homo sapiens	INCY- Human transporter and ion channel-6 (TRICH-6) protein.	1339	92
1620	AAM25877	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1392.	1284	100
1620	gi12856598	Mus musculus	putative	382	64
1621	AAB94278	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14707.	1465	100
1621	gi13543448	Homo sapiens	hypothetical protein FLJ12750, clone MGC:4691 IMAGE:3533384, mRNA, complete cds.	1465	100
1621	gi10434428	Homo sapiens	cDNA FLJ12750 fis, clone NT2RP2001168, weakly similar to VERPROLIN.	1465	100
1622	AAV02669	Homo sapiens	HUMA- Human secreted protein encoded by gene 20 clone HMKAH10.	288	100
1622	gi7510421	Caenorhabditis elegans	hypothetical protein Y6G8.1 - Caenorhabditis elegans >	66	37
1623	AAB95393	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17745.	795	100
1623	gi10435217	Homo sapiens	cDNA FLJ13265 fis, clone OVARC1000937.	795	100
1623	AAB57019	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1597.	275	91
1624	AAG892	Homo sapiens	GEST Human secreted protein,	936	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	81		SEQ ID NO: 401.		
1624	AAU15932	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 885.	936	100
1624	gi10439193	Homo sapiens	cDNA: FLJ22700 fis, clone HSI12073.	936	100
1625	gi12857964	Mus musculus	putative	1533	86
1625	gi10241397	Homo sapiens	Human DNA sequence from clone RP3-336K20 on chromosome 6 Contains parts of 2 genes for novel proteins, ESTs, STSs and GSSs, complete sequence.	964	100
1625	gi16552303	Homo sapiens	cDNA FLJ32234 fis, clone PLACE6004687.	721	97
1626	gi10438158	Homo sapiens	cDNA: FLJ21940 fis, clone HEP04512.	3307	99
1626	AAAG73712	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4476.	455	98
1626	gi10440614	Oryza sativa	putative ATP-dependent RNA helicase	452	32
1627	gi10437837	Homo sapiens	cDNA: FLJ21687 fis, clone COL09385.	1466	100
1627	gi6180178	Homo sapiens	transcription factor IGHM enhancer 3, JM11 protein, JM4 protein, JM5 protein, T54 protein, JM10 protein, A4 differentiation-dependent protein, triple LIM domain protein 6, and synaptophysin genes, complete cds; and L-type calcium channel alpha-1 subunit gene, partial cds, complete sequence.	1182	98
1627	ABB11561	Homo sapiens	HYSE- Human JM10 protein homologue, SEQ ID NO:1931.	947	100
1628	AAM25227	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:742.	2069	100
1628	AAB85362	Homo sapiens	INCY- Human phosphatase (PP) (clone ID 2522707CD1).	2021	100
1628	gi15080505	Homo sapiens	Similar to RIKEN cDNA 5730568A12 gene, clone MGC:17651 IMAGE:3857480, mRNA, complete cds.	1907	100
1629	gi6562845	Rattus norvegicus	type A/B hnRNP p40	1661	91
1629	gi3377279	Rattus norvegicus	AlF-C1	1654	91
1629	gi1814274	Homo sapiens	Human apobec-1 binding protein 1 mRNA, complete cds.	1631	92
1630	AAY36083	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 468.	430	98
1630	AAG00583	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4664.	414	100
1630	gi358259	Cnncidophorus	NADH dehydrogenase subunit 4	66	37

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	4	tigris			
1631	AAB10284	Homo sapiens	GEMV Human fetal placenta protein fragment AC175.2i.	852	94
1631	gi3983463	Homo sapiens	microfibril-associated glycoprotein 2 (MAGP2) gene, exon 10 and complete cds.	852	94
1631	gi13543486	Homo sapiens	Microfibril-associated glycoprotein-2, clone MGC:14490 IMAGE:4247343, mRNA, complete cds.	852	94
1632	gi16198456	Homo sapiens	Similar to RIKEN cDNA 0610040E02 gene, clone MGC:17973 IMAGE:3919892, mRNA, complete cds.	1339	99
1632	AAY02661	Homo sapiens	HUMA- Human secreted protein encoded by gene 12 clone HFTCU19.	1142	99
1632	gi16740689	Mus musculus	RIKEN cDNA 0610040E02 gene	1059	77
1633	AAB53094	Homo sapiens	GETH Human angiogenesis-associated protein PRO826, SEQ ID NO:158.	510	100
1633	AAB50916	Homo sapiens	GETH Human PRO826 protein.	510	100
1633	AAB65204	Homo sapiens	GETH Human PRO826 (UNQ467) protein sequence SEQ ID NO:201.	510	100
1634	AAB53094	Homo sapiens	GETH Human angiogenesis-associated protein PRO826, SEQ ID NO:158.	413	85
1634	AAB50916	Homo sapiens	GETH Human PRO826 protein.	413	85
1634	AAB65204	Homo sapiens	GETH Human PRO826 (UNQ467) protein sequence SEQ ID NO:201.	413	85
1635	gi10439008	Homo sapiens	cDNA: FLJ22573 fis, clone HSI02387.	1937	100
1635	AAM87876	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:15469.	187.	45
1635	gi1334643	Xenopus laevis	APEG precursor protein	93	37
1636	gi10439008	Homo sapiens	cDNA: FLJ22573 fis, clone HSI02387.	578	100
1636	AAM87876	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:15469.	155	94
1636	gi3941365	Homo sapiens	I-REL gene, exon 12 and complete cds.	79	37
1637	AAM93871	Homo sapiens	HEL1- Human polypeptide, SEQ ID NO: 3980.	3761	99
1637	gi14331131	Homo sapiens	scinderin mRNA, complete cds.	3749	99
1637	AAR80481	Homo sapiens	NAKA/Recombinant human adseverin.	3527	92
1638	gi10436970	Homo sapiens	cDNA: FLJ20991 fis, clone CAB02103.	239	67

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1638	gi6460240	Deinococcus radiodurans	DNA-binding response regulator	86	37
1638	gi16754877	Cyprinus carpio	Smad4 type4	79	27
1639	AAY84901	Homo sapiens	INCY- A human proliferation and apoptosis related protein.	2821	95
1639	gi12053225	Homo sapiens	mRNA; cDNA DKFZp434P2235 (from clone DKFZp434F2235); complete cds.	2806	95
1639	gi37330	Homo sapiens	H.sapiens mRNA for tre oncogene (clone 210).	2053	78
1640	AAY84901	Homo sapiens	INCY- A human proliferation and apoptosis related protein.	2846	96
1640	gi12053225	Homo sapiens	mRNA; cDNA DKFZp434P2235 (from clone DKFZp434F2235); complete cds.	2834	95
1640	gi37330	Homo sapiens	H.sapiens mRNA for tre oncogene (clone 210).	2050	77
1641	gi13936285	Mus musculus	TRH4	1522	77
1641	gi12845540	Mus musculus	putative	1520	77
1641	AAU00782	Homo sapiens	INCY- Human apoptosis protein, APOP-2.	1345	98
1642	gi17225331	Homo sapiens	MY0876G05 protein (MY876) mRNA, complete cds.	1209	100
1642	gi12002042	Homo sapiens	brain my048 protein mRNA, complete cds.	1209	100
1642	gi17646146	Homo sapiens	B lymphocyte activation-related protein mRNA, complete cds.	911	78
1643	gi10437307	Homo sapiens	cDNA: FLJ21240 fis, clone COL01132.	2090	100
1643	AAB74730	Homo sapiens	INCY- Human membrane associated protein MEMAP-36.	856	42
1643	AAY94906	Homo sapiens	GEMMY Human secreted protein clone r6649_3 protein sequence SEQ ID NO:18.	853	42
1644	AAY14448	Homo sapiens	HUMA- Human secreted protein encoded by gene 38 clone HFGAH44.	316	100
1645	gi10334443	Homo sapiens	Human DNA sequence from clone RP11-291L22 on chromosome 10 Contains the 3' end of the HSD17B7 (hydroxysteroid (17-beta) dehydrogenase 7) gene, part of a gene similar to CDC10 (cell division cycle 10, S. cerevisiae, homolog), part of a novel gene, a novel pseudogene, STSs, GSSs and a CpG Island, complete sequence.	256	100
1645	gi560623	human, fetal lung, mRNA, 2314 nt. [Homo	hCDC10=CDC10 homolog	236	72

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens			
1645	gi2864606	Mus musculus	CDC10	236	72
1646	AAB18969	Homo sapiens	INCY- Amino acid sequence of a human transmembrane protein.	1707	95
1646	AAB49504	Homo sapiens	HUMA- Clone HNTIMH27.	1370	94
1646	gi16307593	Mus musculus	RIKEN cDNA 2210021G21 gene	1325	90
1647	gi10047231	Homo sapiens	mRNA for KIAA1578 protein, partial cds.	2083	95
1647	gi6841194	Homo sapiens	HSPC272	281	81
1647	gi10800375	Caenorhabditis elegans	Hypothetical protein Y67D8C.5	192	21
1648	gi14349355	Homo sapiens	hypothetical protein FLJ23323, clone MGC:14873 IMAGE:3948222, mRNA, complete cds.	1771	100
1648	gi10439969	Homo sapiens	cDNA: FLJ23323 fis, clone HEP12456.	1771	100
1648	gi12852502	Mus musculus	putative	1540	65
1649	AAY41360	Homo sapiens	HUMA- Human secreted protein encoded by gene 53 clone HUPAD75.	490	100
1649	AAM24406	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1931.	335	100
1649	AAY41470	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 53.	84	100
1650	AAY41360	Homo sapiens	HUMA- Human secreted protein encoded by gene 53 clone HUPAD75.	267	63
1650	AAM24406	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1931.	199	90
1650	AAY41470	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 53.	84	100
1651	gi16550312	Homo sapiens	cDNA FLJ30993 fis, clone HLUNG100064, weakly similar to KARYOGAMY PROTEIN KAR4.	2449	100
1651	gi16306892	Homo sapiens	clone MGC:2902 IMAGE:3010654, mRNA, complete cds.	2449	100
1651	gi13938595	Homo sapiens	Similar to CG7818 gene product, clone MGC:4531 IMAGE:3010654, mRNA, complete cds.	2449	100
1652	AAG67158	Homo sapiens	MILL- Amino acid sequence of a human 20685 transporter polypeptide.	1586	100
1652	gi13274122	Homo sapiens	Human DNA sequence from clone RP1-55C23 on chromosome 6q22.3-23.3 Contains the VNN1 and VNN2	1586	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			genes for vanin 1 and 2, the gene for vanin 3 (VNN3), a HLF (hepatic leukemia factor) pseudogene, a CCNG1 (cyclin G1) pseudogene, the 3' part of a novel gene, ESTs, GSSs, and STSS, complete sequence.		
1652	gi15795817	Homo sapiens	unnamed protein product	1586	100
1653	AAB54163	Homo sapiens	HUMA- Human pancreatic cancer antigen protein sequence SEQ ID NO:615.	322	63
1653	gi180886	Homo sapiens	Human colipase mRNA, complete cds.	322	63
1653	gi17389764	Homo sapiens	colipase, pancreatic, clone MGC:23801 IMAGE:4251084, mRNA, complete cds.	322	63
1654	AAU17198	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 763.	1359	100
1654	gi15987493	Homo sapiens	tumor endothelial marker 6 (TEM6) mRNA, complete cds.	1359	100
1654	gi14325770	Homo sapiens	mRNA for thyroid specific PTB domain protein, complete cds.	1359	100
1655	AAG01118	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5199.	360	100
1655	gi15418966	Xenopus laevis	annexin 4	66	44
1656	gi10437031	Homo sapiens	cDNA: FLJ21034 fls, clone CAE09073.	1403	100
1656	gi16307159	Mus musculus	Unknown (protein for IMAGE:3493084)	1060	50
1656	gi13879551	Mus musculus	Unknown (protein for IMAGE:3709003)	1060	50
1657	AAM80047	Homo sapiens	HYSE- Human protein SEQ ID NO 3693.	820	91
1657	AAM79063	Homo sapiens	HYSE- Human protein SEQ ID NO 1725.	820	89
1657	ABB12144	Homo sapiens	HYSE- Human HSPP-29 protein homologue, SEQ ID NO:2514.	820	91
1658	gi5813823	Homo sapiens	SUJ1 isolog mRNA, complete cds.	470	86
1658	gi450281	Homo sapiens	sui1 iso1 mRNA, complete cds.	470	86
1658	gi14250520	Homo sapiens	putative translation initiation factor, clone MGC:15684 IMAGE:3350981, mRNA, complete cds.	470	86
1659	gi16877187	Homo sapiens	clone MGC:17299 IMAGE:3845811, mRNA, complete cds.	1094	100
1659	AAY12952	Homo sapiens	HUMA- Amino acid sequence of a human secreted peptide.	362	98
1659	gi2394376	Cercopithecus aethiops	thromboxane A2 receptor; TBXA2R	94	29
1660	gi10438946	Homo sapiens	cDNA: FLJ22527 fls, clone HRC12820.	1017	97
1660	gi165493	Homo sapiens	cDNA FLJ30149 fls, clone	665	90

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	12		BRACE2000280, weakly similar to MNN4 PROTEIN.		
1660	gi7299207	Drosophila melanogaster	CG16789 gene product	519	42
1661	AAB95572	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18221.	2231	100
1661	gi10435927	Homo sapiens	cDNA FLJ13798 fis, clone THYRO1000124.	2231	100
1661	gi12851620	Mus musculus	putative	1745	77
1662	AAY48256	Homo sapiens	META- Human prostate cancer-associated protein 42.	242	75
1662	gi3820857	Euglena spirogyra	maturase-like protein	81	30
1662	gi11559649	Leuconostoc mesenteroides	dextranucrase Dsrb742	79	39
1663	gi14017783	Homo sapiens	mRNA for KIAA1783 protein, partial cds.	2287	100
1663	AAU17193	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 758.	1725	100
1663	AAY57895	Homo sapiens	INCY- Human transmembrane protein HTMPN-19.	1634	100
1664	gi10436992	Homo sapiens	cDNA: FLJ21007 fis, clone CAE03871.	3436	100
1664	gi13784943	Mus musculus	Unknown (protein for MGC:11761)	2930	84
1664	gi17862868	Drosophila melanogaster	RE01471p	308	28
1665	AAG93318	Homo sapiens	NISC- Human protein HP10505.	465	100
1665	gi16306868	Homo sapiens	mitochondrial ribosomal protein S21, clone MGC:2680 IMAGE:2819715, mRNA, complete cds.	465	100
1665	gi13620911	Homo sapiens	MRPS21 mRNA for mitochondrial ribosomal protein S21, complete cds.	465	100
1666	AAU27652	Homo sapiens	ZYMO Human protein AFP213641.	1484	100
1666	gi15862470	Homo sapiens	unnamed protein product	1484	100
1666	AAE06071	Homo sapiens	HUMA- Human gene 31 encoded secreted protein HBJLF01, SEQ ID NO:133.	1284	100
1667	AAW78132	Homo sapiens	HUMA- Human secreted protein encoded by gene 7 clone HPEBD85.	246	100
1668	AAM50213	Homo sapiens	CURA- Human interleukin-11-like AMF7 C-terminal polypeptide.	2219	99
1668	gi16548921	Homo sapiens	unnamed protein product	2219	99
1668	AAB94803	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15937.	1358	100
1669	gi12053365	Homo sapiens	mRNA; cDNA DKFZp586O0222 (from clone	2747	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			DKFZp586O0222; complete cds.		
1669	gi9368538	Homo sapiens	mRNA full length insert cDNA clone EUROIMAGE 1987170.	2687	98
1669	gi9965905	Mus musculus	syembryn	2383	86
1670	gi10440280	Homo sapiens	cDNA: FLJ23554 fis, clone LNG09359.	3757	100
1670	gi12855247	Mus musculus	putative	1339	64
1670	gi14602609	Homo sapiens	hypothetical protein FLJ23554, clone MGC:14866 IMAGE:3946091, mRNA, complete cds.	1236	99
1671	gi10441732	Homo sapiens	leucine-rich repeat-containing G protein-coupled receptor 6 (LGR6) mRNA, partial cds.	4286	99
1671	gi17382894	Homo sapiens	unnamed protein product	3899	93
1671	gi17382882	Mus musculus	unnamed protein product	3477	84
1672	AAB9418	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14362.	1936	100
1672	gi10434108	Homo sapiens	cDNA FLJ12552 fis, clone NTZRM4000712, moderately similar to Homo sapiens ubiquitin hydrolyzing enzyme 1 (UBH1) mRNA.	1936	100
1672	AAB95806	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18793.	1930	99
1673	AAB48293	Homo sapiens	UYVA Human ZF5 protein.	1407	80
1673	gi6456114	Mus musculus	F-box protein FBX16	1407	80
1673	ABB15590	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 4247.	831	90
1674	gi10047163	Homo sapiens	mRNA for KIAA1549 protein, partial cds.	7563	100
1674	AAM79157	Homo sapiens	HYSE- Human protein SEQ ID NO 1819.	948	28
1674	AAM80141	Homo sapiens	HYSE- Human protein SEQ ID NO 3787.	941	30
1675	gi16768654	Drosophila melanogaster	HL01494p	911	39
1675	gi7292299	Drosophila melanogaster	CG1271 gene product	888	38
1675	gi4981995	Thermotoga maritima	glycerol kinase	846	38
1676	gi12852837	Mus musculus	putative	892	69
1676	gi3228237	Homo sapiens	UHS KerB gene.	871	71
1676	gi200962	Mus musculus	serine 1 ultra high sulfur protein	827	65
1677	gi17390182	Homo sapiens	clone IMAGE:4797244, mRNA, partial cds.	1694	99
1677	AAY761	Homo sapiens	HUMA- Human secreted protein	863	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	77		encoded by gene 54.		
1677	AAY04306	Homo sapiens	HUMA- Human secreted protein encoded by gene 14.	328	93
1678	gi5689417	Homo sapiens	mRNA for KIAA1040 protein, partial cds.	2793	99
1678	gi10728660	Drosophila melanogaster	CG8683 gene product	2736	48
1678	AAY02367	Homo sapiens	ONQY Polypeptide identified by the signal sequence trap method.	2663	99
1679	gi10439964	Homo sapiens	cDNA: FLJ23320 fis, clone HEP12381.	3605	99
1679	AAG74499	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5263.	623	95
1679	gi12830679	Drosophila helvetica	putative transposase	220	24
1680	gi10438277	Homo sapiens	cDNA: FLJ22028 fis, clone HEP08589.	2454	100
1680	AAB73681	Homo sapiens	INCY- Human oxidoreductase protein ORP-14.	2337	100
1680	gi7298659	Drosophila melanogaster	CG10721 gene product	882	43
1681	AAB94393	Homo sapiens	HELL- Human protein sequence SEQ ID NO:14937.	3231	99
1681	gi10434765	Homo sapiens	cDNA FLJ12973 fis, clone NT2RP2006023.	3231	99
1681	gi12860450	Mus musculus	putative	1177	64
1682	gi10437748	Homo sapiens	cDNA: FLJ21615 fis, clone COL07393.	876	100
1682	gi13195151	Homo sapiens	transcription factor T2P (T2P) mRNA, complete cds.	362	47
1682	gi10241461	Homo sapiens	Human DNA sequence from clone RP5-1121G12 on chromosome 20 Contains the 3' end of a gene encoding the hepatocellular carcinoma-associated antigen 58 (HCA58), the SCAND1 gene encoding the domain-containing 1 protein, a novel gene, 2 CpG islands, ESTs, STSs and GSSs, complete sequence.	362	47
1683	AAB43900	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1345.	483	87
1683	gi1684917	Homo sapiens	Bruton's tyrosine kinase (BTK), alpha-D-galactosidase A (GLA), L44-like ribosomal protein (L44L) and FTP3 (FTP3) genes, complete cds.	483	87
1683	gi12804705	Homo sapiens	ribosomal protein L44, clone MGC:2064 IMAGE:3353669, mRNA, complete cds.	483	87
1684	AAY07931	Homo sapiens	HUMA- Human secreted protein fragment encoded from gene 80.	213	100
1684	gi593200	Mus musculus	neuronal apoptosis inhibitory	68	50

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	3		protein-rs6		
1684	gi5932008	Mus musculus	neuronal apoptosis inhibitory protein	68	50
1685	gi14030407	Mus musculus	keratin-associated protein 16.4	380	77
1685	gi14030409	Mus musculus	keratin-associated protein 16.5	309	64
1685	gi14030401	Mus musculus	keratin-associated protein 16.1	302	67
1686	gi15341794	Homo sapiens	hypothetical protein FLJ12787, clone MGC:16870 IMAGE:3048487, mRNA, complete cds.	1433	100
1686	AAB95182	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:17250.	1426	99
1686	gi10434481	Homo sapiens	cDNA FLJ12787 fls, clone NT2RP2001943.	1426	99
1687	AAAY38401	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 16.	230	88
1687	AAB25687	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 23 SEQ ID NO:76.	66	38
1688	gi12744921	Homo sapiens	tehering factor SEC34 (SEC34) mRNA, complete cds.	4223	100
1688	gi14549669	Homo sapiens	vesicle docking protein SEC34 mRNA, complete cds.	4212	99
1688	gi15291537	Drosophila melanogaster	GH25768p	1691	43
1689	AAB90746	Homo sapiens	GEMY Human DF989_3 protein sequence SEQ ID 192.	545	97
1689	AAW64471	Homo sapiens	GEMY Human secreted protein from clone DF989_3.	545	97
1689	gi2829302	Homo sapiens	mRNA for Ets1, complete cds.	74	37
1690	AAW13658	Homo sapiens	UYMC- Human cytidine deaminase.	657	87
1690	gi598149	Homo sapiens	cytidine deaminase (CDA) mRNA, complete cds.	657	87
1690	gi4321793	Homo sapiens	cytidine deaminase gene, exon 4 and complete cds.	657	87
1691	gi12698079	Homo sapiens	mRNA for KIAA1767 protein, partial cds.	4013	99
1691	AAM25578	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1093.	3955	99
1691	AAE06186	Homo sapiens	HUMA- Human gene 58 encoded secreted protein fragment, SEQ ID NO:248.	3521	99
1692	gi2559010	Homo sapiens	chaperonin containing t-complex polypeptide 1, eta subunit (Cct1) mRNA, complete cds.	890	100
1692	gi14198388	Mus musculus	chaperonin subunit 7 (eta)	879	98
1692	gi468504	Mus musculus	CCTeta, eta subunit of the chaperonin containing TCP-1 (CCT)	879	98
1693	gi16552036	Homo sapiens	cDNA FLJ32028 fls, clone NTONG1000257.	918	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1693	AAB74768	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 13 SEQ ID NO:77.	592	99
1693	AAB74745	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 13 SEQ ID NO:54.	592	99
1694	gi10047157	Homo sapiens	mRNA for KIAA1546 protein, partial cds.	3652	100
1694	gi12697897	Homo sapiens	mRNA for KIAA1676 protein, partial cds.	613	39
1694	gi7292283	Drosophila melanogaster	CG2083 gene product	534	35
1695	AAG00078	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4159.	164	80
1695	gi2370202	Homo sapiens	mRNA for procollagen alpha 2(V).	164	80
1695	gi179698	Homo sapiens	Human collagen type V alpha-2 mRNA, 5' end.	164	80
1696	gi16552596	Homo sapiens	cDNA FLJ32466 fis, clone SKNMC2000065.	2609	99
1696	gi14017827	Homo sapiens	mRNA for KIAA1805 protein, partial cds.	2609	99
1696	gi15207987	Macaca fascicularis	hypothetical protein	2588	99
1697	AAB92879	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11475.	639	73
1697	AAM41435	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6366.	639	73
1697	AAM39649	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2794.	639	73
1698	AAM25487	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1002.	586	100
1698	AAG03667	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7748.	582	99
1698	gi12850050	Mus musculus	putative	557	93
1699	AAE04774	Homo sapiens	INCY- Human vesicle trafficking protein-17 (VETR-17) protein.	748	100
1699	AAB41637	Homo sapiens	CURA- Human ORFX ORF1401 polypeptide sequence SEQ ID NO:2802.	748	100
1699	gi3319953	Homo sapiens	mRNA for TOM1 protein.	638	82
1700	gi14140100	Homo sapiens	OTT gene for one twenty two proteins (spliced and unspliced forms).	4797	99
1700	AAB95111	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17093.	4779	99
1700	gi14161369	Homo sapiens	putative RNA-binding motif protein 15 short form (RBM15) mRNA, complete cds, alternatively spliced.	4779	99
1701	AAB93879	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13792.	2228	100
1701	AAG66710	Homo sapiens	BIOD- Human cell growth inhibition protein 48.	2228	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1701	gi14495627	Homo sapiens	clone MGC:15047 IMAGE:3535485, mRNA, complete cds.	2228	100
1702	AAE11780	Homo sapiens	INCY- Human kinase (PKIN)-14 protein.	4186	100
1702	gi14041817	Homo sapiens	gklp mRNA for kinase-like protein splice variant 1, complete cds.	4186	100
1702	AAB65679	Homo sapiens	SUGE- Novel protein kinase, SEQ ID NO: 207.	4158	97
1703	AAU12253	Homo sapiens	GETH Human PRO5774 polypeptide sequence.	440	74
1703	AAY30734	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	258	96
1703	gi1181669	Saccharomyces cerevisiae	Tel2p	75	24
1704	gi10439762	Homo sapiens	cDNA: FLJ23164 fis, clone LNG09764.	3205	100
1704	gi10440312	Homo sapiens	cDNA: FLJ23577 fis, clone LNG12640.	1755	98
1704	gi12698085	Homo sapiens	mRNA for KIAA1770 protein, partial cds.	1614	99
1705	AAM41806	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6737.	1194	77
1705	AAM40020	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3165.	1194	77
1705	ABB12231	Homo sapiens	HYSE- Human novel protein, SEQ ID NO:2601.	1194	77
1706	gi15277565	Mus musculus	RIKEN cDNA 2510039O18 gene	3091	91
1706	gi12846932	Mus musculus	putative	3088	91
1706	AAB43028	Homo sapiens	CURA- Human ORFX ORF2792 polypeptide sequence SEQ ID NO:5584.	2246	95
1707	gi16041136	Macaca fascicularis	hypothetical protein	702	92
1707	AAB65216	Homo sapiens	GETH Human PRO1004 (UNQ488) protein sequence SEQ ID NO:227.	569	92
1707	AAY66693	Homo sapiens	GETH Membrane-bound protein PRO1004.	569	92
1708	AAB95636	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18369.	2429	100
1708	gi10436357	Homo sapiens	cDNA FLJ14009 fis, clone Y79AA1002431, weakly similar to TRANSDUCIN-LIKE ENHANCER PROTEIN 2.	2429	100
1708	gi5030439	Homo sapiens	chromosome 19, cosmid R26610, complete sequence.	1569	80
1709	gi15705143	Mus musculus	suppressor of cytokine signalling 4	2071	86
1709	AAM00759	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 122.	1712	100
1709	AAM00872	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 348.	1215	99
1710	gi104387	Homo sapiens	cDNA: FLJ22408 fis, clone	1934	100

SEQ ID	HIT ID	Species	Description	S score	Percent identity
	85		HRC08416.		
1710	AAM39917	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3062.	837	45
1710	AAM41703	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6634.	836	45
1711	gi0437428	Homo sapiens	cDNA: FLJ21343 fis, clone COL02679.	1836	100
1711	gi12053145	Homo sapiens	mRNA; cDNA DKFZp434A0926 (from clone DKFZp434A0926); complete cds.	1388	99
1711	gi5689537	Homo sapiens	mRNA for KIAA1100 protein, complete cds.	1261	68
1712	AAG03254	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7335.	438	89
1712	gi7300383	Drosophila melanogaster	CG7671 gene product	366	27
1712	AAB9451	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15312.	309	100
1713	gi1449568	Homo sapiens	hypothetical protein FLJ12687, clone MGC:15791 IMAGE:3504468, mRNA, complete cds.	2639	99
1713	AAB94241	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14627.	2631	99
1713	gi10434333	Homo sapiens	cDNA FLJ12687 fis, clone NT2RM4002532, weakly similar to PROTEIN HOM1.	2631	99
1714	AAC62351_aa1	Homo sapiens	CELL- Nucleotide sequence of lysophosphatidic acid acyltransferase-beta.	834	100
1714	AAA39292_aa1	Homo sapiens	CELL- Human lysophosphatidic acid acyltransferase beta encoding cDNA.	834	100
1714	AAB30623	Homo sapiens	CELL- Amino acid sequence of lysophosphatidic acid acyltransferase-beta.	834	100
1715	AAB95847	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18896.	675	100
1715	gi10436763	Homo sapiens	cDNA FLJ14326 fis, clone PLACE4000247.	675	100
1715	gi16552900	Homo sapiens	cDNA FLJ32711 fis, clone TEST12000707, weakly similar to DOUBLESEX PROTEIN, MALE-SPECIFIC.	90	35
1716	AAB53094	Homo sapiens	GETH Human angiogenesis-associated protein PRO826, SEQ ID NO:158.	278	100
1716	AAB50916	Homo sapiens	GETH Human PRO826 protein.	278	100
1716	AAB65204	Homo sapiens	GETH Human PRO826 (UNQ467) protein sequence SEQ ID NO:201.	278	100
1717	AAB94915	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16349.	961	100
1717	gi14602623	Homo sapiens	hypothetical protein FLJ11526, clone MGC:15059	961	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			IMAGE:3937610, mRNA, complete cds.		
1717	gi10432797	Homo sapiens	cDNA FLJ11526 fls, clone HEMBA1002555, weakly similar to Homo sapiens mSin3A associated polypeptide p30 mRNA.	961	100
1718	AAY12439	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID NO:470.	453	94
1718	gi13477183	Homo sapiens	Similar to hypothetical protein FLJ20859, clone MGC:12940 IMAGE:282127, mRNA, complete cds.	453	94
1718	gi12830810	Homo sapiens	false p73 target protein gene, complete cds.	453	94
1719	gi15928965	Homo sapiens	hypothetical protein FLJ11354, clone MGC:22961 IMAGE:4865798, mRNA, complete cds.	3522	100
1719	AAB93708	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:13299.	3514	99
1719	gi10432595	Homo sapiens	cDNA FLJ11354 fls, clone HEMBA1000129, weakly similar to HYPOTHETICAL HELICASE C8A4.08C IN CHROMOSOME 1.	3514	99
1720	gi14717396	Homo sapiens	potassium-dependent Na/Ca exchanger NCKX3 (SLC24A3) mRNA, partial cds.	3108	97
1720	gi12597441	Mus musculus	K ⁺ -dependent Na/Ca exchanger	3027	94
1720	gi12000397	Rattus norvegicus	potassium-dependent sodium-calcium exchanger NCKX3	3025	94
1721	gi15072454	Mus musculus	von Willbrand factor A-related protein	1614	72
1721	AAB42581	Homo sapiens	CURA- Human ORFX ORF2345 polypeptide sequence SEQ ID NO:4690.	1358	93
1721	AAB88340	Homo sapiens	HEL1- Human membrane or secretory protein clone PSEC0053.	1288	98
1722	AAE03822	Homo sapiens	HUMA- Human gene 5 encoded secreted protein HETKL27, SEQ ID NO: 68.	935	100
1722	AAB82597	Homo sapiens	HUMA- Human transmembrane protein encoded by cDNA clone HNALE36.	935	100
1722	gi15706437	Homo sapiens	clone MGC:17366 IMAGE:3860009, mRNA, complete cds.	935	100
1723	AAG66503	Homo sapiens	BIOD- Human ATP-dependent helicase 31.	1441	100
1723	AAM25780	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1295.	1441	100
1723	gi13676356	Homo sapiens	clone MGC:2679 IMAGE:2819663, mRNA, complete cds.	1434	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1724	AAG673 94	Homo sapiens	SUGE- Amino acid sequence of human protein kinase SGK269.	2322	100
1724	gi104371 81	Homo sapiens	cDNA: FLJ21140 fis, clone CAS07548.	1730	100
1724	AAG673 93	Homo sapiens	SUGE- Amino acid sequence of human protein kinase SGK223.	952	46
1725	AAF2449 8 aa1	Homo sapiens	GEST Human PG-3 coding sequence.	4362	99
1725	AAB354 01	Homo sapiens	GEST Human PG-3.	4355	99
1725	gi133968 64	Homo sapiens	unnamed protein product	4355	99
1726	AAB427 84	Homo sapiens	CURA- Human ORFX ORF2548 polypeptide sequence SEQ ID NO:5096.	817	99
1726	gi122248 87	Homo sapiens	mRNA; cDNA DKFZp547H027 (from clone DKFZp547H027); complete cds.	817	99
1726	gi104384 59	Homo sapiens	cDNA: FLJ22174 fis, clone HRC00767.	817	99
1727	gi167405 66	Homo sapiens	Similar to hypothetical protein FLJ13087, clone MGC:15009 IMAGE3:3536735, mRNA, complete cds.	1854	100
1727	AAB952 97	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:17525.	807	95
1727	gi104349 41	Homo sapiens	cDNA FLJ13087 fis, clone NT2RP302009.	807	95
1728	AAB940 75	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:14267.	3506	97
1728	AAM939 95	Homo sapiens	HEL1- Human stomach cancer expressed polypeptide SEQ ID NO 59.	3506	97
1728	gi140421 45	Homo sapiens	cDNA FLJ14550 fis, clone NT2RM2001696.	3506	97
1729	AAG008 97	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4978.	392	93
1729	AAG008 98	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4979.	130	100
1729	AAB425 97	Homo sapiens	CURA- Human ORFX ORF2361 polypeptide sequence SEQ ID NO:4722.	130	100
1730	ABB1161 3	Homo sapiens	HYSE- Human sorting nexin 7 homologue, SEQ ID NO:1983.	2341	100
1730	AAG741 74	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4938.	2309	99
1730	gi488424 1	Homo sapiens	mRNA; cDNA DKFZp564F052 (from clone DKFZp564F052); partial cds.	2148	99
1731	gi134456 60	Homo sapiens	MP19 (LIM2) mRNA, complete cds, alternatively spliced.	933	100
1731	gi111775 46	Homo sapiens	LIM2 (LIM2) and natural killer group 7 (NKG7) genes, complete cds.	933	100
1731	gi134456 58	Homo sapiens	MP19ins (LIM2) mRNA, complete cds, alternatively	880	80

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			spliced.		
1732	AAB40591	Homo sapiens	CURA: Human ORFX ORF355 polypeptide sequence SEQ ID NO:710.	1056	100
1732	gi15617458	Oryctolagus cuniculus	Rab11 family interacting protein	869	48
1732	AAY29488	Homo sapiens	CORI: Human lung tumour protein LT86-7 predicted amino acid sequence.	557	87
1733	gi10437750	Homo sapiens	cDNA: FLJ21616 fis, clone COL07477.	1680	99
1733	gi14349360	Homo sapiens	Similar to hypothetical protein FLJ21616, clone MGC:14941 IMAGE:3947903, mRNA, complete cds.	1443	99
1733	gi12805473	Mus musculus	Unknown (protein for IMAGE:3490304)	1410	97
1734	gi10437750	Homo sapiens	cDNA: FLJ21616 fis, clone COL07477.	1645	92
1734	gi14349360	Homo sapiens	Similar to hypothetical protein FLJ21616, clone MGC:14941 IMAGE:3947903, mRNA, complete cds.	1565	99
1734	gi12805473	Mus musculus	Unknown (protein for IMAGE:3490304)	1412	96
1735	AAB95036	Homo sapiens	HELL: Human protein sequence SEQ ID NO:16791.	863	100
1735	gi10433448	Homo sapiens	cDNA FLJ12060 fis, clone HEMBB1002142.	863	100
1735	gi15341904	Homo sapiens	clone MGC:21051 IMAGE:4476886, mRNA, complete cds.	751	99
1736	gi14043783	Homo sapiens	clone MGC:14256 IMAGE:4129368, mRNA, complete cds.	2232	100
1736	gi10436857	Homo sapiens	cDNA: FLJ20897 fis, clone ADKA03573.	2232	100
1736	gi12653785	Homo sapiens	clone IMAGE:3349601, mRNA, partial cds.	1783	99
1737	gi14249850	Homo sapiens	clone MGC:15062 IMAGE:2959567, mRNA, complete cds.	1535	99
1737	AAM79539	Homo sapiens	HYSE: Human protein SEQ ID NO 3185.	1523	53
1737	AAM78555	Homo sapiens	HYSE: Human protein SEQ ID NO 1217.	1523	53
1738	gi13183338	Homo sapiens	calneuron 1 (CALN1) mRNA, complete cds.	881	100
1738	gi13183340	Mus musculus	calneuron 1	880	99
1738	gi7670344	Mus musculus	unnamed protein product	880	99
1739	gi4995927	Xenopus laevis	p33 ringo	545	49
1739	gi10280285	Xenopus sp.	unnamed protein product	545	49
1739	gi446879	Xenopus laevis	speedy protein	540	48

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	0				
1740	gi15281553	Homo sapiens	solute carrier family 12 member 8 (SLC12A8) gene, partial cds.	1905	80
1740	gi17224940	Mus musculus	cation-chloride cotransporter	1761	74
1740	gi10439794	Homo sapiens	cDNA: FLJ23188 fis, clone LNG12038.	1613	99
1741	gi6330163	Homo sapiens	mRNA for KIAA1161 protein, partial cds.	1137	99
1741	gi7301333	Drosophila melanogaster	CG11909 gene product	485	41
1741	gi64404	Torpedo californica	4-acetamido-4'-isothiocyanostilbene-2, 2'-disulphonic acid-binding protein	415	38
1742	AAM67857	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28163.	1553	100
1742	AAM55471	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 27576.	1553	100
1742	gi15620849	Homo sapiens	mRNA for KIAA1895 protein, partial cds.	1553	100
1743	gi2852640	Homo sapiens	clone 23856 unknown mRNA, partial cds.	942	99
1743	gi13325281	Homo sapiens	hypothetical protein MGC2683, clone MGC:4313 IMAGE:2819900, mRNA, complete cds.	700	100
1743	gi12654485	Homo sapiens	clone MGC:2683 IMAGE:2819900, mRNA, complete cds.	700	100
1744	gi13938307	Homo sapiens	clone MGC:15626 IMAGE:3343642, mRNA, complete cds.	526	62
1744	AAB90765	Homo sapiens	NOJ1/ Human shear stress-response protein SEQ ID NO: 30.	524	64
1744	AAB62159	Homo sapiens	NEUR- Human arginine-rich protein.	524	64
1745	gi7243207	Homo sapiens	mRNA for KIAA1413 protein, partial cds.	7273	99
1745	AAB93057	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11862.	3014	99
1745	gi7022861	Homo sapiens	cDNA FLJ10680 fis, clone NT2RP2006573, weakly similar to 2',3'-CYCLIC NUCLEOTIDE 3'-PHOSPHODIESTERASE (EC 3.1.4.37).	3014	99
1746	gi7243207	Homo sapiens	mRNA for KIAA1413 protein, partial cds.	7245	98
1746	AAB93057	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11862.	3014	99
1746	gi7022861	Homo sapiens	cDNA FLJ10680 fis, clone NT2RP2006573, weakly similar to 2',3'-CYCLIC NUCLEOTIDE 3'-PHOSPHODIESTERASE (EC 3.1.4.37).	3014	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1747	AAE10330	Homo sapiens	INCY- Human transporter and ion channel-7 (TRICH-7) protein.	1520	100
1747	AAM39422	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2567.	694	48
1747	AAM79397	Homo sapiens	HYSE- Human protein SEQ ID NO 3043.	694	48
1748	gi10439744	Homo sapiens	cDNA: FLJ23151 fls, clone LNG09417.	2362	100
1748	gi12860456	Mus musculus	putative	1732	70
1748	gi8954063	Arabidopsis thaliana	Contains similarity to a transposable element Tip100 protein for transposase from Ipomoea purpurea gb 4063769 and is a member of the transmembrane 4 family PF00335.	308	24
1749	gi17529987	Homo sapiens	oxysterol-binding protein-like protein OSBPL3 (OSBPL3) mRNA, complete cds.	4671	100
1749	gi17389382	Homo sapiens	oxysterol binding protein-like 3, clone MGC:21526 IMAGE:3909164, mRNA, complete cds.	4671	100
1749	gi10880973	Homo sapiens	oxysterol binding protein-related protein 3 (ORP3) mRNA, complete cds.	4671	100
1750	AAB94879	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16094.	2813	100
1750	gi10436338	Homo sapiens	cDNA FLJ13998 fls, clone Y79AA1002229, weakly similar to DNA CROSS-LINK REPAIR PROTEIN PSO2/SNM1.	2813	100
1750	AAB94319	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14797.	2330	99
1751	AAB19390	Homo sapiens	LEX1- Amino acid sequence of a human lipoxygenase protein.	3836	100
1751	gi13378170	Homo sapiens	partial ALOXE3 gene for arachidonate lipoxygenase 3, exons 1 to 4B (and joined CDS).	3836	100
1751	gi10441004	Homo sapiens	epidermal lipoxygenase (ALOXE3) mRNA, complete cds.	3830	99
1752	AAM93241	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2671.	664	95
1752	gi15030270	Homo sapiens	clone MGC:9889 IMAGE:3868330, mRNA, complete cds.	664	95
1752	AAO10285	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 24177.	486	78
1753	gi456191	Homo sapiens	H.sapiens mRNA for rho GDP-dissociation Inhibitor I.	818	99
1753	gi337395	Homo sapiens	Human GDP dissociation inhibitor mRNA, complete cds.	818	99
1753	gi285979	Homo sapiens	Human rho GDI mRNA, complete cds.	818	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1754	AAQ435 49_aa1	Homo sapiens	HARD Gamma subunit of human retinal cGMP phosphodiesterase DNA.	472	100
1754	AAR931 17	Homo sapiens	HARD cGMP-phosphodiesterase gamma-subunit.	472	100
1754	AAR384 84	Homo sapiens	HARD Gamma subunit of human retinal cGMP phosphodiesterase.	472	100
1755	AAY194 46	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	590	98
1755	AAY195 99	Homo sapiens	HUMA- SEQ ID NO 317 from WO9922243.	590	98
1755	AAY196 02	Homo sapiens	HUMA- SEQ ID NO 320 from WO9922243.	137	100
1756	gi104373 93	Homo sapiens	cDNA: FLJ21313 fls, clone COL02176.	2197	99
1756	gi142503 21	Homo sapiens	hypothetical protein FLJ21313, clone MGC:16820 IMAGE:4148772, mRNA, complete cds.	2193	99
1756	gi128585 26	Mus musculus	putative	1936	86
1757	gi741435 1	Homo sapiens	Hox1.8 gene for homeobox protein.	478	100
1757	gi278967 2	Homo sapiens	homeobox protein A10 (HOXA10) gene, complete cds.	478	100
1757	gi155592 35	Homo sapiens	clone MGC:12859 IMAGE:4107013, mRNA, complete cds.	478	100
1758	gi189772	Homo sapiens	Human prostaglandin D2 synthase gene, exons 2 through 6 and complete cds.	872	90
1758	gi135435 68	Homo sapiens	prostaglandin D2 synthase (21kD, brain), clone MGC:14559 IMAGE:4294999, mRNA, complete cds.	872	90
1758	gi129638 79	Homo sapiens	prostaglandin D synthase mRNA, complete cds.	872	90
1759	gi298250 8	Homo sapiens	mRNA for TCR beta chain, specific for Mage 3/HLA-A2.	1296	92
1759	gi300292 5	Homo sapiens	T cell receptor beta chain (TCRBV13S1-TCRB12S1) mRNA, complete cds.	1286	92
1759	gi36733	Homo sapiens	H.sapiens mRNA for T-cell antigen receptor beta-chain.	1047	75
1760	AAM257 28	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1243.	667	99
1760	gi155290 64	Homo sapiens	sorting nexin 14 (SNX14) mRNA, complete cds.	667	99
1760	gi134772 73	Homo sapiens	clone MGC:13217 IMAGE:3959086, mRNA, complete cds.	667	99
1761	AAB948 43	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16018.	1761	100
1761	gi104362 67	Homo sapiens	cDNA FLJ13955 fls, clone Y79AA100117.	1761	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1761	gi16359295	Mus musculus	Similar to hypothetical protein FLJ13955	1681	85
1762	AAE02058	Homo sapiens	HUMA- Human four disulfide core domain (FDCD)-containing protein.	1200	84
1762	gi12835376	Mus musculus	putative	918	68
1762	gi12655452	Homo sapiens	mRNA for keratin associated protein 4.7 (KRTAP4.7 gene).	892	68
1763	AAG67485	Homo sapiens	LEX1- Amino acid sequence of a human transporter protein.	2391	99
1763	AAE10333	Homo sapiens	INCY- Human transporter and ion channel-10 (TRICH-10) protein.	2368	96
1763	gi12718201	Homo sapiens	Human DNA sequence from clone RP11-305P22 on chromosome 20 Contains ESTs, STSs, GSSs and 7 CpG islands. Contains three novel genes and a novel gene for a helix-loop-helix DNA binding protein, complete sequence.	2243	100
1764	gi15778948	Homo sapiens	Similar to thiamine pyrophosphokinase, clone MGC:14885 IMAGE:3622116, mRNA, complete cds.	717	100
1764	gi12667203	Homo sapiens	thiamine pyrophosphokinase (TPK1) mRNA, complete cds.	717	100
1764	gi12248915	Homo sapiens	hTPK1 mRNA for thiamin pyrophosphokinase, complete cds.	717	100
1765	gi10438831	Homo sapiens	cDNA: FLJ22439 fis, clone HRC09236.	2525	99
1765	AAB42237	Homo sapiens	CURA- Human ORFX ORF2001 polypeptide sequence SEQ ID NO:4002.	2084	99
1765	gi13559284	Homo sapiens	Human DNA sequence from clone RP5-117516 on chromosome 20. Contains the 3' end of the gene for Ras inhibitor JC265 (Ras association (RalGDS/AF-6) domain containing protein), the 5' end of the gene encoding N-terminal acetyltransferase complex and 1 subunit, ESTs, STSs, GSSs and two CpG islands, complete sequence.	829	36
1766	gi9588428	Homo sapiens	Human DNA sequence from clone RP5-1024N4 on chromosome 1p32.1-33. Contains the gene for a novel Sodium:solute symporter family member similar to SLC5A1 (SGLT1), a pseudogene similar to part of butyrophilin family members, a novel gene, ESTs, STSs, GSSs and a putative CpG	2858	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			island, complete sequence.		
1766	gi529056	Homo sapiens	Na ⁺ /glucose cotransporter (SGLT1) gene, exon 15 and complete cds.	1955	55
1766	gi3646043	Homo sapiens	Human DNA sequence from PAC I27L4 on chromosome 22. Contains last exon (15) of the SLC5A1 (SGLT1) gene for solute carrier family 5 (sodium/glucose cotransporter) member 1 (High Affinity Sodium-Glucose Cotransporter). Contains ESTs and STSs, complete sequence.	1955	55
1767	AAB73485	Homo sapiens	MILL- Human aminopeptidase 22196.	3657	99
1767	gi12583612	Homo sapiens	mRNA for neurolysin.	3657	99
1767	gi13922467	Homo sapiens	unnamed protein product	3657	99
1768	gi10122138	Rattus norvegicus	SynGAP-a	6651	99
1768	gi2935448	Rattus norvegicus	synaptic ras GTPase-activating protein p135 SynGAP	6634	99
1768	gi6624587	Homo sapiens	Human DNA sequence from clone RP4-570F3 on chromosome 6. Contains the 5' end of the gene for the ortholog of the rat synaptic ras GTPase-activating protein p135 SynGAP, gene LOC51596 for divalent cation tolerant protein CUTA or brain acetylcholinesterase putative membrane anchor, the PHF1 gene for PHD finger protein 1, the KNSL2 gene for kinesin-like protein 2, the gene for a novel protein similar to ribosomal protein L12 (RPL12) and the gene for a novel protein similar to lysophospholipase II (LYPLA2). Contains ESTs, STSs, GSSs and four CpG islands, complete sequence.	6194	100
1769	AAB95863	Homo sapiens	HELL- Human protein sequence SEQ ID NO:18931.	1040	100
1769	gi17389283	Homo sapiens	hypothetical protein FLJ14346, clone MGC:21027 IMAGE:4415420, mRNA, complete cds.	1040	100
1769	gi10436791	Homo sapiens	cDNA FLJ14346 fis, clone THYRO1001320.	1040	100
1770	AAB94517	Homo sapiens	HELL- Human protein sequence SEQ ID NO:15235.	2593	99
1770	gi10435122	Homo sapiens	cDNA FLJ13203 fis, clone NT2RP3004504, highly similar to M.musculus mRNA for CPFB	2593	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			protein.		
1770	gi12659216	Homo sapiens	cytoplasmic polyadenylation element-binding protein short form (CPEB1) mRNA, complete cds.	2583	98
1771	AAB48059	Homo sapiens	INCY- Human extracellular signaling molecule (EXCS) (ID 1493630CD1).	432	97
1771	AAAY36090	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 475.	420	95
1771	AAAY11768	Homo sapiens	GEST Human 5' EST secreted protein SEQ ID No: 368.	257	95
1772	AAB93075	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11902.	2936	99
1772	gi14042415	Homo sapiens	cDNA FLJ14710 fis, clone NT2RP3000632, weakly similar to ZINC FINGER PROTEIN 84.	2936	99
1772	gi16551429	Homo sapiens	cDNA FLJ31551 fis, clone NT2R12001083, moderately similar to ZINC FINGER PROTEIN 84.	1813	63
1773	AAAY76843	Homo sapiens	INCY- Human proton ATPase subunit (HPAS) protein sequence.	356	100
1773	AAAY88590	Homo sapiens	SATO/ Human tumour specific anti/gen amino acid sequence.	356	100
1773	AAW64534	Homo sapiens	SAGA Human fibrosarcoma cell line HT-1080 clone HP00442 protein.	356	100
1774	gi12407385	Homo sapiens	tripartite motif protein TRIM5 isoform gamma (TRIM5) mRNA, complete cds; alternatively spliced.	1818	99
1774	gi12407387	Homo sapiens	tripartite motif protein TRIM5 isoform delta (TRIM5) mRNA, complete cds; alternatively spliced.	1559	99
1774	gi12407383	Homo sapiens	tripartite motif protein TRIM5 isoform beta (TRIM5) mRNA, complete cds; alternatively spliced.	1557	100
1775	gi14250601	Homo sapiens	hypothetical protein FLJ22056, clone MGC:3045 IMAGE:3343082, mRNA, complete cds.	2600	99
1775	gi10438315	Homo sapiens	cDNA: FLJ22056 fis, clone HEP09916.	1747	100
1775	gi7295293	Drosophila melanogaster	CG8633 gene product	597	31
1776	gi10437059	Homo sapiens	cDNA: FLJ21054 fis, clone CAS00538.	736	100
1776	gi11558099	Mus musculus	syncoilin	653	89
1776	gi12835405	Mus musculus	putative	552	84
1777	gi163076	Homo sapiens	hypothetical protein FLJ12270,	3045	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	08		clone MGC:10176 IMAGE:3908004, mRNA, complete cds.		
1777	gi15620905	Homo sapiens	mRNA for KIAA1923 protein, partial cds.	3042	99
1777	AAB93945	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:13963.	2190	100
1778	gi7959819	Homo sapiens	PRO1430	299	100
1778	AAO07840	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 21732.	75	54
1778	AAO08870	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 22762.	73	43
1779	AAG62910	Homo sapiens	KLEE/ Amino acid sequence of a human xylosyltransferase (XT) isoform XT-II.	4625	100
1779	gi11322270	Homo sapiens	mRNA for xylosyltransferase II (XT-II gene).	4625	100
1779	gi15209653	Homo sapiens	human XT-II	4625	100
1780	gi9929973	Macaca fascicularis	hypothetical protein	1778	96
1780	AAG78740	Homo sapiens	BODE- Human transcriptional elongation factor IIS 24.	1162	99
1780	ABB11220	Homo sapiens	HYSE- Human TF11Sh homologue, SEQ ID NO:1590.	653	100
1781	gi15277846	Homo sapiens	Similar to hypothetical protein FLJ21522, clone MGC:16817 IMAGE:3853503, mRNA, complete cds.	3122	99
1781	AAB64372	Homo sapiens	INCY- Amino acid sequence of human intracellular signalling molecule INTRA4.	3100	99
1781	gi10437638	Homo sapiens	cDNA: FLJ21522 fls, clone COL05884.	2892	94
1782	AAB94625	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:15483.	2688	100
1782	gi10435387	Homo sapiens	cDNA FLJ13386 fls, clone PLACE1001104, weakly similar to MYOSIN HEAVY CHAIN, NON-MUSCLE.	2688	100
1782	gi16551877	Homo sapiens	cDNA FLJ31903 fls, clone N12RP7004260, weakly similar to MYOSIN HEAVY CHAIN, NONMUSCLE TYPE B.	2429	92
1783	gi10568112	Homo sapiens	ALR-like protein mRNA, complete cds.	17050	100
1783	gi5630077	Homo sapiens	PAC clone RP5-98107 from 7q34-q36, complete sequence.	9606	100
1783	AAB42230	Homo sapiens	CURA- Human ORFX ORF1994 polypeptide sequence SEQ ID NO:3988.	9583	99
1784	AAB68876	Homo sapiens	INCY- Human RECAP polypeptide, SEQ ID NO: 6.	1895	85
1784	gi10438174	Homo sapiens	cDNA: FLJ21952 fls, clone HEP04970.	1895	85
1784	gi169242	Homo sapiens	hypothetical protein FLJ21952,	1862	84

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	96		clone MGC:2790 IMAGE:2960984, mRNA, complete cds.		
1785	AAU160 26	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 979.	1433	100
1785	gi140178 23	Homo sapiens	mRNA for KIAA1803 protein, partial cds.	1433	100
1785	AAG780 55	Homo sapiens	GEAT Human zinc finger domain DNA binding protein S 1-3.	1168	99
1786	AAG004 05	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4486.	397	98
1786	AAM906 02	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:18195.	215	56
1786	AAM411 14	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6045.	84	34
1787	AAG020 95	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6176.	302	100
1787	gi665103 7	Mus musculus domesticus	similar to RNA binding protein	222	53
1787	gi128478 83	Mus musculus	putative	222	53
1788	gi979845 2	Homo sapiens	mRNA for putative capacitative calcium channel (trp7 gene).	4470	100
1788	gi532685 4	Mus musculus	receptor-activated calcium channel	4392	98
1788	gi229590 3	Homo sapiens	Human putative calcium influx channel (htrp3) mRNA, complete cds.	3529	81
1789	AAG023 37	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6418.	377	98
1789	AAM008 75	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 351.	72	31
1789	gi166669 2	Mus musculus	alpha-NAC, muscle-specific form gp220	72	34
1790	AAB733 81	Homo sapiens	NANF- Human gas vesicle protein homologue hGvpT-b.	2838	99
1790	gi120055 09	Homo sapiens	HT025 mRNA, complete cds.	2838	99
1790	gi173914 58	Homo sapiens	clone MGC:2462 IMAGE:2964737, mRNA, complete cds.	1699	99
1791	AAY994 38	Homo sapiens	GETH Human PRO1555 (UNQ763) amino acid sequence SEQ ID NO:338.	1300	100
1791	AAB240 37	Homo sapiens	GETH Human PRO1555 protein sequence SEQ ID NO:49.	1300	100
1791	gi126542 33	Homo sapiens	Similar to hypothetical protein, clone 1-2, clone MGC:5442 IMAGE:3449979, mRNA, complete cds.	1300	100
1792	AAB638 60	Homo sapiens	LUDW- Human prostate cancer associated antigen protein sequence SEQ ID NO:1222.	297	50
1792	AAM767 71	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein	272	51

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			SEQ ID NO: 37077.		
1792	AAM63951	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 36056.	272	51
1793	AAB94249	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14645.	799	99
1793	gi10434356	Homo sapiens	cDNA FLJ12700 fis, clone NT2RP1000721.	799	99
1793	ABB12422	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 261.	528	99
1794	AAB65189	Homo sapiens	GETH Human PRO1013 (UNQ496) protein sequence SEQ ID NO:158.	1858	99
1794	AAB87536	Homo sapiens	GETH Human PRO1013.	1858	99
1794	AAY66666	Homo sapiens	GETH Membrane-bound protein PRO1013.	1858	99
1795	AAB65189	Homo sapiens	GETH Human PRO1013 (UNQ496) protein sequence SEQ ID NO:158.	1655	93
1795	AAB87536	Homo sapiens	GETH Human PRO1013.	1655	93
1795	AAY66666	Homo sapiens	GETH Membrane-bound protein PRO1013.	1655	93
1796	gi7770263	Homo sapiens	PRO3077	620	100
1796	gi15891632	Agrobacterium tumefaciens	AGR_L_3035p	67	31
1797	AAY36233	Homo sapiens	HUMA- Human secreted protein encoded by gene 10.	302	96
1797	gi298307	Rattus sp.	beta 3-adrenergic receptor; beta 3-AR	83	38
1797	gi241216	Rattus sp.	beta 3-adrenergic receptor	83	38
1798	ABB12037	Homo sapiens	HYSE- Human ribosomal protein L31 homologue, SEQ ID NO:2407.	341	100
1798	AAG03894	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7975.	341	100
1798	AAB43707	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1152.	341	100
1799	AAE10995	Homo sapiens	INCY- Human lipid metabolism enzyme-4 (LME-4) protein.	2242	99
1799	AAB41989	Homo sapiens	CURA- Human ORFX ORF1753 polypeptide sequence SEQ ID NO:3506.	2224	98
1799	AAB94007	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14123.	1212	99
1800	AAY76194	Homo sapiens	HUMA- Human secreted protein encoded by gene 71.	296	98
1800	AAY13196	Homo sapiens	GEST Human secreted protein encoded by 3' EST SEQ ID NO: 210.	291	96
1800	AAY19471	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	148	96
1801	ABB1242	Homo sapiens	HYSE- Human bone marrow	2841	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	2		expressed protein SEQ ID NO: 261.		
1801	gi3638956	Homo sapiens	PAC clone RP4-751H13 from 7q35-qter, complete sequence.	2830	100
1801	gi7670496	Mus musculus	unnamed protein product	2352	84
1802	AAU12382	Homo sapiens	GETH Human PRO792 polypeptide sequence.	1332	86
1802	AAB24416	Homo sapiens	GETH Human PRO792 protein sequence SEQ ID NO:155.	1332	86
1802	AAB24055	Homo sapiens	GETH Human PRO792 protein sequence SEQ ID NO:31.	1332	86
1803	AAU12382	Homo sapiens	GETH Human PRO792 polypeptide sequence.	1590	100
1803	AAB24416	Homo sapiens	GETH Human PRO792 protein sequence SEQ ID NO:155.	1590	100
1803	AAB24055	Homo sapiens	GETH Human PRO792 protein sequence SEQ ID NO:31.	1590	100
1804	AAB43713	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1158.	414	98
1804	AAR11490	Homo sapiens	CALI- Tissue-plastin.	414	98
1804	gi339848	Homo sapiens	Human T-plastin mRNA, 5' end.	414	98
1805	AAI19456	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	307	90
1805	gi13882500	Mycobacterium tuberculosis CDC1551	conserved hypothetical transmembrane protein	70	32
1805	gi1550714	Mycobacterium tuberculosis H37Rv	hypothetical protein Rv2673	70	32
1806	gi16552850	Homo sapiens	cDNA FLJ32676 fis, clone TEST11000168, weakly similar to PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT SDS22.	1265	100
1806	gi12859936	Mus musculus	putative	962	78
1806	AAM38877	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2022.	200	45
1807	AAY41390	Homo sapiens	HUMA- Human secreted protein encoded by gene 83 clone HRAAB15.	813	100
1807	AAM39990	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3135.	581	44
1807	AAM38999	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2144.	581	44
1808	gi12836197	Mus musculus	putative	2154	75
1808	AAM82398	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:9991.	750	90
1808	AAG04069	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8150.	491	100
1809	AAB95252	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17419.	3112	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1809	gi10434729	Homo sapiens	cDNA FLJ12949 fis, clone NT2RP2005336, weakly similar to TRICHOHYALIN.	3112	99
1809	gi12804075	Homo sapiens	hypothetical protein FLJ12949, clone MGC:11261 IMAGE:3942403, mRNA, complete cds.	1786	100
1810	AAY19587	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	160	96
1811	gi15620881	Homo sapiens	mRNA for KIAA1911 protein, partial cds.	1153	100
1811	AAM91641	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:19234.	326	50
1811	gi12845802	Mus musculus	putative	309	58
1812	gi10439911	Homo sapiens	cDNA: FLJ23282 fis, clone HEP07626.	2677	100
1812	AAY99653	Homo sapiens	INCY- Human GTPase associated protein-4.	2324	100
1812	gi1016712	Rattus norvegicus	Fos-related antigen	2060	84
1813	AAY60152	Homo sapiens	META- Human endometrium tumour EST encoded protein 212.	379	100
1813	AAY60151	Homo sapiens	META- Human endometrium tumour EST encoded protein 211.	83	73
1813	AAB92984	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11704.	71	32
1814	gi14602995	Homo sapiens	hypothetical protein FLJ23375, clone MGC:16634 IMAGE:4121449, mRNA, complete cds.	2299	100
1814	gi10440040	Homo sapiens	cDNA: FLJ23375 fis, clone HEP16206.	2294	99
1814	AAB42973	Homo sapiens	CURA- Human ORFX ORF2737 polypeptide sequence SEQ ID NO:5474.	725	98
1815	AAM78722	Homo sapiens	HYSE- Human protein SEQ ID NO 1384.	1808	99
1815	gi15053987	Homo sapiens	c-Mpl binding protein mRNA, complete cds.	1439	100
1815	AAE10199	Homo sapiens	HYSE- Human bone marrow derived contig polypeptide, SEQ ID NO: 64.	1231	85
1816	gi10440474	Homo sapiens	mRNA for FLJ00074 protein, partial cds.	1001	100
1816	gi7020969	Homo sapiens	cDNA FLJ20703 fis, clone KAI11965.	546	63
1816	gi13325142	Homo sapiens	DKFZP586I2223 protein, clone MGC:10840 IMAGE:3616057, mRNA, complete cds.	535	58
1817	gi10047249	Homo sapiens	mRNA for KIAA1587 protein, partial cds.	5002	100
1817	gi13676486	Macaca fascicularis	hypothetical protein	4641	92

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1817	gi12659140	Mus musculus	mage-e1	2474	67
1818	AAV72596	Homo sapiens	ZYMO Human cytokine alpha protein-27 (Zalpha27)	3538	99
1818	gi12710297	Homo sapiens	unnamed protein product	3538	99
1818	gi10437588	Homo sapiens	cDNA: FLJ21478 fis, clone COL05012.	2793	100
1819	gi12838732	Mus musculus	putative	1060	91
1819	gi6572215	Homo sapiens	Human DNA sequence from clone RP1-37E16 on chromosome 22. Contains the 3' part of the gene for a novel VHS domain containing protein similar to predicted worm and human proteins, the SH3BP1 gene for SH3-domain binding protein 1, the gene for a novel protein similar to nitrophenylphosphatases from various organisms, the LGALS1 gene for soluble galactoside-binding lectin 1, a novel gene and the gene for a novel protein similar to mouse RBP3 (P116 Rho-interacting protein) and rat RB109, complete sequence.	647	46
1819	gi12653107	Homo sapiens	hypothetical protein dJ37E16.5, clone MGC:8472 IMAGE:2821743, mRNA, complete cds.	647	46
1820	AAB73690	Homo sapiens	INCY- Human oxidoreductase protein ORP-23.	2502	100
1820	gi10438222	Homo sapiens	cDNA: FLJ21988 fis, clone HEP06320.	2502	100
1820	gi14336719	Homo sapiens	16p13.3 sequence section 3 of 8.	2431	90
1821	gi3724105	Homo sapiens	hHa4 gene for keratin type I.	2029	99
1821	gi3724101	Homo sapiens	hHa3-I gene for keratin type I.	1798	90
1821	gi12852606	Mus musculus	putative	1796	87
1822	gi3228237	Homo sapiens	UHS KerB gene.	1079	88
1822	AAM79404	Homo sapiens	HYSE- Human protein SEQ ID NO 3050.	877	81
1822	gi200962	Mus musculus	serine 1 ultra high sulfur protein	867	70
1823	AAB86364	Homo sapiens	MEMO- Human ceramidase K2 protein.	1459	100
1823	AAB18986	Homo sapiens	INCY- Amino acid sequence of a human transmembrane protein.	1459	100
1823	gi14669436	Homo sapiens	alkaline phytoecerdamide (APHC) mRNA, complete cds.	1459	100
1824	AAW61601	Homo sapiens	INCY- Human metallothionein HMBP-L.	381	98

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1824	AAB57183	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1761.	363	84
1824	AAO13869	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 27761.	362	81
1825	gi15928896	Homo sapiens	Similar to synaptotagmin-like 4, clone MGC:17313 IMAGE:3908307, mRNA, complete cds.	3496	100
1825	gi17385944	Rattus norvegicus	granuphilin A	3225	91
1825	gi5926736	Mus musculus	granuphilin-a	3187	90
1826	gi12653147	Homo sapiens	signal sequence receptor, beta (translocon-associated protein bcta), clone MGC:8566 IMAGE:2822983, mRNA, complete cds.	864	99
1826	gi452757	Homo sapiens	H.sapiens mRNA for TRAP beta subunit.	847	99
1826	gi1736880	Homo sapiens	Human SSR2 mRNA for beta-signal sequence receptor, complete cds.	847	99
1827	AAB42722	Homo sapiens	CURA- Human ORFX ORF2486 polypeptide sequence SEQ ID NO:4972.	2865	100
1827	AAB95263	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:17448.	2864	99
1827	gi10434759	Homo sapiens	cDNA FLJ12969 fis, clone NT2RP2005841, weakly similar to Homo sapiens mRNA for ALEX3.	2864	99
1828	gi7981297	Homo sapiens	Human DNA sequence from clone RP4-534K7 on chromosome 1p31.2-32.3. Contains the PGM1 gene for phosphoglucomutase 1, a novel gene, ESTs, STS, GSSs and a putative CpG island, complete sequence.	3323	100
1828	gi16041846	Homo sapiens	clone MGC:9635 IMAGE:3915942, mRNA, complete cds.	3308	99
1828	gi14017815	Homo sapiens	mRNA for KIAA1799 protein, partial cds.	3186	100
1829	gi9438033	Homo sapiens	ser/arg-rich pre-mRNA splicing factor SR-A1 (SR-A1) gene, complete cds.	6833	100
1829	gi10440402	Homo sapiens	mRNA for FLJ00034 protein, partial cds.	6827	99
1829	gi1438532	Rattus norvegicus	rA1	5019	82
1830	AAG81294	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:106.	1128	100
1830	AAU03589	Homo sapiens	INCY- Human DNA modification protein, DNAMP-4.	1128	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1830	gi17390202	Homo sapiens	Similar to RIKEN cDNA 2510005D08 gene, clone MGC:27120 IMAGE:4793121, mRNA, complete cds.	1128	100
1831	AAY65282	Homo sapiens	GEST Human 5' EST related polypeptide SEQ ID NO:1443, putative	472	95
1831	gi12838627	Mus musculus		444	68
1831	gi6513867	Strongylocentrotus purpuratus	tektin A1	292	40
1832	AAY10837	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	434	100
1832	gi7271606	Fowlpox virus	ORF FPV108 Virion envelope protein	72	28
1832	gi333523	Pigeonpox virus	major envelope antigen	72	28
1833	AAR56166	Homo sapiens	USSH Neuroendocrine tumor dlk.	2003	95
1833	gi15341994	Homo sapiens	clone MGC:17291 IMAGE:4347187, mRNA, complete cds.	2003	95
1833	gi15559310	Homo sapiens	clone MGC:20310 IMAGE:4130556, mRNA, complete cds.	2000	94
1834	AAM51465	Homo sapiens	TAKE Human G protein-coupled receptor protein TGR5.	1712	100
1834	AAY86540	Homo sapiens	HUMA- Human gene 77- encoded protein fragment, SEQ ID NO:457.	679	98
1834	AAY86291	Homo sapiens	HUMA- Human secreted protein HDPRK33, SEQ ID NO:206.	514	98
1835	gi16552598	Homo sapiens	cDNA FLJ32467 fis, clone SKNMC2000097, moderately similar to M.musculus mRNA for protein Htf9C.	2883	99
1835	gi15426494	Homo sapiens	HpaII tiny fragments locus 9C, clone MGC:14943 IMAGE:4054100, mRNA, complete cds.	2517	100
1835	gi10437555	Homo sapiens	cDNA: FLJ21453 fis, clone COL04585.	2517	100
1836	gi7243061	Homo sapiens	mRNA for KIAA1340 protein, partial cds.	2328	99
1836	gi3882143	Homo sapiens	mRNA for KIAA0711 protein, complete cds.	311	28
1836	gi14602992	Homo sapiens	clone MGC:16635 IMAGE:4121528, mRNA, complete cds.	309	29
1837	AAB94727	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15753.	2624	96
1837	gi10435806	Homo sapiens	cDNA FLJ13710 fis, clone PLACE2000373, weakly similar to F-SPONDIN PRECURSOR.	2624	96
1837	gi14250742	Homo sapiens	hypothetical protein FLJ13710, clone MGC:14171 IMAGE:4120678, mRNA, complete cds.	2473	98
1838	AAB199	Homo sapiens	INCY- Human oxidoreductase	1900	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	31		OXR-6.		
1838	gi17044356	Homo sapiens	unnamed protein product	1900	99
1838	AAW67737	Homo sapiens	REGC Human fsh05 gene protein product.	1528	100
1839	gi10440014	Homo sapiens	cDNA: FLJ23356 fis, clone HEP14919.	1859	99
1839	gi15208223	Macaca fascicularis	hypothetical protein	1798	96
1839	gi12853689	Mus musculus	putative	1535	80
1840	gi14250512	Homo sapiens	clone MGC:15468 IMAGE:2966921, mRNA, complete cds.	966	88
1840	gi14043262	Homo sapiens	Similar to RIKEN cDNA 1500026B10 gene, clone MGC:15737 IMAGE:3355622, mRNA, complete cds.	966	88
1840	gi12837754	Mus musculus	putative	710	68
1841	gi17671629	Homo sapiens	Human DNA sequence from clone RP11-145L22 on chromosome 6p21.32-22.2. Contains the gene for myelin/oligodendrocyte glycoprotein (MOG), the gene for a novel KRAB box containing C2H2 type zinc finger protein, ESTs, STSS, GSSs and a CpG island, complete sequence.	2247	90
1841	gi431182	Mus musculus	Zfp-57	674	39
1841	gi14495650	Homo sapiens	zinc finger protein 331; zinc finger protein 463, clone MGC:15739 IMAGE:3355780, mRNA, complete cds.	444	29
1842	gi10439951	Homo sapiens	cDNA: FLJ23311 fis, clone HEP11681.	2175	99
1842	ABB12461	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 300.	259	100
1842	gi5541722	Arabidopsis thaliana	putative protein	218	44
1843	gi9992884	Homo sapiens	vacuolar proton pump 116 kDa accessory subunit (ATP6N1B) mRNA, complete cds, alternatively spliced.	4378	100
1843	AAB60100	Homo sapiens	INCY- Human transport protein TPPT-20.	4331	99
1843	gi16903213	Mus musculus	H-ATPase accessory subunit a4	3809	85
1844	AAM93859	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3953.	3149	100
1844	gi17224957	Homo sapiens	cationic amino acid transporter (SLC7A3) mRNA, complete cds.	3149	100
1844	AAM993	Homo sapiens	HELI- Human polypeptide, SEQ	3142	99

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	85		ID NO: 2970.		
1845	gi12053007	Homo sapiens	mRNA; cDNA DKFZp434D1812 (from clone DKFZp434D1812); complete cds.	6038	99
1845	gi4240195	Homo sapiens	mRNA for KIAA0853 protein, partial cds.	4974	100
1845	AAB67047	Homo sapiens	INCY- Human immune response molecule (IMUN) protein SEQ ID NO: 1.	3777	99
1846	AAB94108	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14340.	2903	100
1846	AAU04557	Homo sapiens	GETH Human Stra6 homologue, PRO10282.	2903	100
1846	gi13560966	Homo sapiens	STRA6 isoform 1 mRNA, complete cds, alternatively spliced.	2903	100
1847	AAB58363	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 701.	309	100
1847	AAY48507	Homo sapiens	META- Human breast tumour-associated protein 52.	308	98
1847	AAM23952	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1477.	294	98
1848	AAM93737	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3705.	929	49
1848	gi170609	Rattus norvegicus	monocarboxylate transporter MCT3	631	35
1848	gi7670446	Mus musculus	unnamed protein product	631	47
1849	AAB95359	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17661.	3130	99
1849	gi10435088	Homo sapiens	cDNA FLJ13181 fis, clone NT2RP3004016, weakly similar to TRANSCRIPTION INTERMEDIARY FACTOR 1-BETA.	3130	99
1849	gi7297357	Drosophila melanogaster	CG8419 gene product	746	29
1850	gi10436813	Homo sapiens	cDNA: FLJ20859 fis, clone ADKA01617.	2426	100
1850	gi13477183	Homo sapiens	Similar to hypothetical protein FLJ20859, clone MGC:12940 IMAGE:2822127, mRNA, complete cds.	2357	98
1850	gi12830810	Homo sapiens	false p73 target protein gene, complete cds.	2239	99
1851	AAM24367	Homo sapiens	HYSE- Human EST encoded protein SEQ ID NO: 1892.	546	100
1851	AAAY27576	Homo sapiens	HUMA- Human secreted protein encoded by gene No. 10.	394	96
1851	gi30478	Homo sapiens	Human mRNA for dopamine beta-hydroxylase type b (EC 1.14.17.1).	68	26
1852	gi16551561	Homo sapiens	cDNA FLJ31657 fis, clone NT2RI2004304, moderately similar to Homo sapiens NY-	2859	92

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			REN-2 antigen mRNA.		
1852	AAB36626	Homo sapiens	INCY- Human FLEXHT-48 protein sequence SEQ ID NO:48.	2056	67
1852	gi12803469	Homo sapiens	high-glucose-regulated protein 8, clone MGC:739 IMAGE:3139250, mRNA, complete cds.	2056	67
1853	gi7768736	Homo sapiens	genomic DNA, chromosome 21q, section 87/105.	4306	99
1853	gi14245729	Homo sapiens	ANKRD3 mRNA for dual-specificity Ser/Thr/Tyr kinase, complete cds.	4003	94
1853	gi9886711	Homo sapiens	mRNA for protein kinase (dik gene).	3999	94
1854	gi10439129	Homo sapiens	cDNA: FLJ22655 fls, clone HSI07590.	960	100
1854	AAG01459	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5540.	668	100
1854	AAU17365	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 930.	523	100
1855	gi10047185	Homo sapiens	mRNA for KIAA1560 protein, partial cds.	3397	100
1855	gi17512495	Mus musculus	glycerol-3-phosphate acyltransferase, mitochondrial	3371	93
1855	gi193367	Mus musculus	glycerol-3-phosphate acyltransferase	3363	93
1856	ABB12236	Homo sapiens	HYSE- Human eppin-1 homologue, SEQ ID NO:2606.	472	100
1856	gi13937334	Homo sapiens	Human DNA sequence from clone RP3-461P17 on chromosome 20q12-13.2. Contains two novel genes, gene HE4 for Major Epididymis-specific protein E4 precursor (Epididymis Secretory protein E4), RPL5 (60S Ribosomal Protein L5), COX6C (Cytochrome C Oxidase subunit VIC) and HSPD1 (HSP60, Mitochondrial Matrix Protein P1 precursor, Heat Shock Protein 60, GROEL, HUCHA60) pseudogenes, the SPINT3 gene for Kunitz type serine protease inhibitor 3 (HKIB9), two genes for novel Kunitz/Bovine pancreatic trypsin inhibitor and WAP-type (Whey Acidic Protein) 'four-disulfide core' domains containing proteins and the gene for Eppin-1, -2 and -3. Contains ESTs, STSs, GSSs and a CpG island, complete sequence.	415	98
1856	gi13591753	Oryctolagus cuniculus	eppin	257	66

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1857	gi12855841	Mus musculus	putative	758	57
1857	gi17736646	Homo sapiens	Human DNA sequence from clone RP3-341D10 on chromosome X Contains a gene for a novel protein, part of the gene for a protein similar to ADP ribosylation factor 3, part of a gene similar to HTF9C and a CpG island, complete sequence.	424	100
1857	AAM38958	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2103.	421	43
1858	gi10439520	Homo sapiens	cDNA: FLJ22973 fls, clone KAT11042.	2289	100
1858	gi11612388	Homo sapiens	zinc finger transcription factor Pegasus mRNA, complete cds.	2279	99
1858	AAU16142	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1095.	1745	98
1859	gi13544026	Homo sapiens	putative zinc finger protein from EUROIMAGE 566589, clone MGC:13109 IMAGE:3959436, mRNA, complete cds.	1481	99
1859	AAB93932	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13929.	1474	99
1859	gi10433647	Homo sapiens	cDNA FLJ12222 fls, clone MAMMA1001105, moderately similar to OVO PROTEIN.	1474	99
1860	AAG78615	Homo sapiens	SHAN- Human zinc finger transcription factor BioZFTF45.	1760	89
1860	gi13623431	Homo sapiens	clone MGC:13132 IMAGE:4124255, mRNA, complete cds.	1753	100
1860	gi13111897	Homo sapiens	Similar to KIAA0414 protein, clone MGC:2629 IMAGE:3503643, mRNA, complete cds.	1753	100
1861	AAB62201	Homo sapiens	RIGE- Cell cycle protein Radh-isoform 1.	3697	99
1861	gi8163804	Mus musculus	putative repair and recombination helicase RAD26L	3215	89
1861	AAB62202	Homo sapiens	RIGE- Cell cycle protein Radh-isoform 2.	2142	99
1862	gi2598565	Mus musculus	rab19	1008	88
1862	AAM78977	Homo sapiens	HYSE- Human protein SEQ ID NO 1639.	603	56
1862	gi7295127	Drosophila melanogaster	Rab-RP3 gene product	523	55
1863	gi17384067	Homo sapiens	Human DNA sequence from clone RP11-146P21 on chromosome 10 Contains the 3'end of a novel gene, a novel gene, the 5'end of the gene for KIAA0608 and a CpG island, complete sequence.	2467	99
1863	gi140399	Cricetulus	hypothetical protein 1-2	2311	92

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	64	griseus			
1863	AAB95196	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:17284.	2183	99
1864	gi12653987	Homo sapiens	cutaneous T-cell lymphoma tumor antigen se70-2, clone MGC:5291 IMAGE:3451565, mRNA, complete cds.	1237	99
1864	gi11385662	Homo sapiens	CTCL tumor antigen se70-2 mRNA, complete cds.	1237	99
1864	gi10436925	Homo sapiens	cDNA: FLJ20957 fis, clone ADSE02053.	1233	99
1865	gi10439911	Homo sapiens	cDNA: FLJ23282 fis, clone HEP07626.	2579	100
1865	AAY99653	Homo sapiens	INCY- Human GTPase associated protein-4.	2544	100
1865	gi1016712	Rattus norvegicus	Fos-related antigen	2256	84
1866	AAB53073	Homo sapiens	GETH Human angiogenesis- associated protein PRO195, SEQ ID NO:46.	331	100
1866	AAB88428	Homo sapiens	HEL1- Human membrane or secretory protein clone PSEC0203.	331	100
1866	AAU12307	Homo sapiens	GETH Human PRO195 polypeptide sequence.	331	100
1867	AAB94197	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:14532.	4854	99
1867	gi10434243	Homo sapiens	cDNA FLJ12634 fis, clone NT2RM4001858, weakly similar to T-BOX CONTAINING PROTEIN TBX6L.	4854	99
1867	gi6692607	Mus musculus	MGA protein	4810	84
1868	AAB95779	Homo sapiens	HEL1- Human protein sequence SEQ ID NO:18726.	1326	100
1868	AAB42878	Homo sapiens	CURA- Human ORFX ORF2642 polypeptide sequence SEQ ID NO:5284.	1326	100
1868	gi10436625	Homo sapiens	cDNA FLJ14220 fis, clone NT2RFP3003828.	1326	100
1869	AAY10823	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	183	100
1869	gi1171584	Plasmodium falciparum	red alga1 chloroplast	72	29
1869	gi630468	Plasmodium falciparum	hypothetical protein 470 - Plasmodium falciparum >	72	29
1870	gi14336713	Homo sapiens	i6p13.3 sequence section 3 of 8.	1867	100
1870	gi5912459	Homo sapiens	Human DNA sequence from clone LA16-380A1 on chromosome 16 Contains two novel genes, ESTs, an STS, GSSs and five putative CpG islands, complete sequence.	1106	100
1870	gi13905252	Mus musculus	Unknown (protein for IMAGE:3601186)	145	28
1871	AAB433	Homo sapiens	CURA- Human ORFX	2463	97

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	82		ORF3146 polypeptide sequence SEQ ID NO:6292.		
1871	AAW734 00	Homo sapiens	HUMA- Human secreted protein encoded by Gene No. 4.	2454	100
1871	AAB583 40	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 678.	2252	90
1872	AAB569 24	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1502.	2048	95
1872	AAY599 93	Homo sapiens	META- Human endometrium tumour EST encoded protein 53.	2048	95
1872	gi476122 3	Homo sapiens	NADP+-dependent isocitrate dehydrogenase (PICD) mRNA, complete cds.	2048	95
1873	AAB643 73	Homo sapiens	INCY- Amino acid sequence of human intracellular signalling molecule INTRA5.	1875	100
1873	AAB949 37	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16430.	1546	99
1873	gi152919 19	Drosophila melanogaster	LD31969p	1237	41
1874	AAB643 73	Homo sapiens	INCY- Amino acid sequence of human intracellular signalling molecule INTRA5.	1836	92
1874	AAB949 37	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16430.	1546	99
1874	gi152919 19	Drosophila melanogaster	LD31969p	1209	39
1875	gi136235 11	Homo sapiens	clone MGC:12921 IMAGE:4129897, mRNA, complete cds.	590	100
1875	gi126982 16	Macaca fascicularis	hypothetical protein	589	99
1875	AAM936 16	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3444.	584	99
1876	gi606368 8	Homo sapiens	ammecr1 gene, exon 1 and joined CDS (alternative transcripts).	968	56
1876	gi513948 2	Homo sapiens	mRNA for AMMECR1 protein.	968	56
1876	gi889465 7	Mus musculus	AMMECR1	964	53
1877	gi104402 18	Homo sapiens	cDNA: FLJ23506 fis, clone LNG03055.	2913	99
1877	AAY733 63	Homo sapiens	INCY- HTRM clone 2762174 protein sequence.	2110	100
1877	AAU162 62	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1215.	1286	98
1878	AAG671 51	Homo sapiens	INCY- Amino acid sequence of a human enzyme.	1689	99
1878	gi128562 10	Mus musculus	putative	1457	85
1878	gi312821 8	Arabidopsis thaliana	putative katanin	874	56
1879	AAB600 93	Homo sapiens	INCY- Human transport protein TPPT-13.	3295	95

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1879	gi17223724	Homo sapiens	sodium/glucose cotransporter KST1 mRNA, complete cds.	3295	95
1879	AAF84023_aa1	Homo sapiens	LEXI- Novel human transporter protein (NHP) encoding DNA.	3289	95
1880	AAB60093	Homo sapiens	INCY- Human transport protein TPPT-13.	3461	99
1880	gi17223724	Homo sapiens	sodium/glucose cotransporter KST1 mRNA, complete cds.	3461	99
1880	AAF84023_aa1	Homo sapiens	LEXI- Novel human transporter protein (NHP) encoding DNA.	3455	99
1881	gi12845475	Mus musculus	putative	1648	70
1881	gi13543307	Homo sapiens	cargo selection protein (mannose 6 phosphate receptor binding protein), clone MGC:11117 IMAGE:3833411, mRNA, complete cds.	725	39
1881	AAAY67240	Homo sapiens	INCY- Human adipophilin-like protein (HALP) amino acid sequence.	724	39
1882	AAW46904	Homo sapiens	ASAH human mutant alanine aminotransferase.	1821	68
1882	gi1763096	Homo sapiens	Human glutamate pyruvate transaminase (GPT) gene, complete cds.	1821	68
1882	gi17390465	Homo sapiens	glutamic-pyruvate transaminase (alanine aminotransferase), clone MGC:17068 IMAGE:4179699, mRNA, complete cds.	1821	68
1883	AAD09495_aa1	Homo sapiens	SUGE- Human SGP003 phosphatase polypeptide encoding DNA.	1161	100
1883	AAE04837	Homo sapiens	SUGE- Human SGP003 phosphatase polypeptide.	1159	99
1883	AAB18667	Homo sapiens	INCY- A human regulator of intracellular phosphorylation.	1021	91
1884	gi3089427	Homo sapiens	SSC6 rearranged T cell receptor beta chain (TCRBV17) gene, complete cds.	1089	69
1884	gi3002927	Homo sapiens	T cell receptor beta chain (TCRBV17S1-TCRBJ1S5) mRNA, complete cds.	1089	69
1884	gi2982508	Homo sapiens	mRNA for TCR beta chain, specific for Mage 3/HLA-A2.	1089	70
1885	gi12858221	Mus musculus	putative	1850	91
1885	gi12314102	Homo sapiens	Human DNA sequence from clone RP3-329L24 on chromosome 6q22.1-22.33 Contains a gene for a novel protein, part of a gene for a hypothetical 23.0 KD protein, part of a gene for a protein (MCM2/3/5 family), ESTs, STSs, GSSs and a CpG island, complete sequence.	1405	100
1885	gi438883	Arabidopsis	putative DNA replication	726	41

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	2	thaliana	licensing factor		
1886	gi14714600	Homo sapiens	clone IMAGE:3354344, mRNA, partial cds.	3001	99
1886	AAR94386	Homo sapiens	NEWE- Human neural cell protein marker RR/B.	673	29
1886	gi3769362	Homo sapiens	ectoderm-neural cortex-1 protein (ENC-1) mRNA, complete cds.	673	29
1887	gi10047239	Homo sapiens	mRNA for KIAA1582 protein, partial cds.	7698	96
1887	gi16551820	Homo sapiens	cDNA FLJ31859 fis, clone NT2RP7001231.	3076	100
1887	gi7959181	Homo sapiens	mRNA for KIAA1460 protein, partial cds.	2840	45
1888	gi10047239	Homo sapiens	mRNA for KIAA1582 protein, partial cds.	7800	99
1888	gi16551820	Homo sapiens	cDNA FLJ31859 fis, clone NT2RP7001231.	2835	93
1888	gi7959181	Homo sapiens	mRNA for KIAA1460 protein, partial cds.	2744	44
1889	gi10047239	Homo sapiens	mRNA for KIAA1582 protein, partial cds.	7372	95
1889	gi7959181	Homo sapiens	mRNA for KIAA1460 protein, partial cds.	2618	44
1889	gi16551820	Homo sapiens	cDNA FLJ31859 fis, clone NT2RP7001231.	2354	81
1890	gi16307285	Homo sapiens	clone IMAGE:3877337, mRNA, partial cds.	1627	98
1890	gi15208051	Macaca fascicularis	hypothetical protein	1417	55
1890	AAY94918	Homo sapiens	GEMY Human secreted protein clone dd504_18 protein sequence SEQ ID NO:42.	1247	63
1891	gi2780414	Homo sapiens	hBACH mRNA for brain acyl-CoA hydrolase, complete cds.	1742	100
1891	gi1692433	Homo sapiens	cytosolic acyl coenzyme A thioester hydrolase, clone MGC:1126 IMAGE:3507488, mRNA, complete cds.	1742	100
1891	AAW74896	Homo sapiens	HUMA- Human secreted protein encoded by gene 169 clone HPTTU11.	1715	98
1892	AAB68874	Homo sapiens	INCY- Human RECAP polypeptide, SEQ ID NO: 4.	1266	100
1892	AAY94890	Homo sapiens	PROT- Human protein clone HP02798.	1266	100
1892	gi14424518	Homo sapiens	Similar to RIKEN cDNA I190004A01 gene, clone MGC:13153 IMAGE:4302257, mRNA, complete cds.	1266	100
1893	gi14595019	Homo sapiens	mRNA for keratin 6 lrs (KRT6LRS gene).	843	77
1893	gi6092075	Mus musculus	type II cytokeratin	836	74
1893	gi12834535	Mus musculus	putative	836	74
1894	gi14029153	Homo sapiens	microtubule-associated protein GLFND mRNA, complete cds.	2565	100

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1894	gi13111907	Homo sapiens	clone MGC:3213 IMAGE:3502614, mRNA, complete cds.	2565	100
1894	gi13447759	Homo sapiens	fibronectin type 3 and SPRY domain-containing protein 1 (FSD1) mRNA, complete cds.	2562	99
1895	gi14424591	Homo sapiens	hypothetical protein FLJ22127, clone MGC:14926 IMAGE:4123948, mRNA, complete cds.	2565	100
1895	gi10438086	Homo sapiens	cDNA: FLJ21886 fis, clone HEP03022.	2555	99
1895	gi10438400	Homo sapiens	cDNA: FLJ22127 fis, clone HEP19530.	1566	100
1896	gi15150306	Homo sapiens	glycogenin-interacting protein 3 mRNA, complete cds.	1571	100
1896	gi15150304	Homo sapiens	glycogenin-interacting protein 2 mRNA, complete cds, alternatively spliced.	1571	100
1896	gi15150298	Homo sapiens	glycogenin-interacting protein 1 mRNA, complete cds.	1571	100

TABLE 3

SEQ ID NO:	Database entry ID	Description	Results*
950	BL00972	Ubiquitin carboxyl-terminal hydrolases family 2 proteins.	BL00972A 11.93 2.500e-20 267-285 BL00972D 22.55 5.179e-17 828-853 BL00972E 20.72 8.650e-13 855-877 BL00972C 16.48 7.120e-11 411-426 BL00972B 9.45 7.923e-10 353-363
950	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 8.000e-10 2-17
950	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308A 5.90 7.671e-09 5-20 PR00308A 5.90 9.471e-09 4-19
951	BL00018	EF-hand calcium-binding domain proteins.	BL00018 7.41 5.696e-09 331-344
951	PR00180	CELLULAR RETINALDEHYDE-BINDING PROTEIN SIGNATURE	PR00180C 10.92 8.821e-09 70-92
952	BL00018	EF-hand calcium-binding domain proteins.	BL00018 7.41 5.696e-09 331-344
952	PR00180	CELLULAR RETINALDEHYDE-BINDING PROTEIN SIGNATURE	PR00180C 10.92 8.821e-09 70-92
953	PR00401	SH2 DOMAIN SIGNATURE	PR00401B 12.94 1.000e-08 340-351
954	PR00401	SH2 DOMAIN SIGNATURE	PR00401B 12.94 1.000e-08 367-378
955	BL00625	Regulator of chromosome condensation (RCC1) proteins.	BL00625A 16.21 7.787e-16 308-337 BL00625A 16.21 7.369e-15 190-219 BL00625B 17.69 1.514e-13 302-336 BL00625B 17.69 2.286e-13 184-218 BL00625B 17.69 3.957e-13 132-166 BL00625A 16.21 5.690e-13 138-167 BL00625A 16.21 5.731e-11 360-389 BL00625B 17.69 3.333e-10 354-388

SEQ ID NO:	Database entry ID	Description	Results*
955	PR00633	CHROMOSOME CONDENSATION REGULATOR RCC1 SIGNATURE	PR00633A 9.32 6.143e-09 202-219 PR00633H 15.10 6.268e-09 196-218 PR00633F 10.03 6.354e-09 373-388 PR00633G 13.71 7.556e-09 190-209
957	BL00120	Lipases, serine proteins.	BL00120B 11.37 9.486e-12 166-181
957	PR00111	ALPHA/BETA HYDROLASE FOLD SIGNATURE	PR00111B 10.61 1.176e-09 170-184
963	BL00183	Ubiquitin-conjugating enzymes proteins.	BL00183 28.97 1.329e-10 45-93
966	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 6.034e-09 262-277
967	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 6.034e-09 74-89
968	BL00790	Receptor tyrosine kinase class V proteins.	BL00790D 12.41 8.297e-09 804-829
969	BL00790	Receptor tyrosine kinase class V proteins.	BL00790D 12.41 8.297e-09 878-903
971	PR00380	KINESIN HEAVY CHAIN SIGNATURE	PR00380D 9.93 2.080e-22 321-343 PR00380A 14.18 1.486e-21 79-101 PR00380B 12.64 6.571e-18 217-235 PR00380C 13.18 6.927e-13 269-288
971	BL00411	Kinesin motor domain proteins.	BL00411H 15.66 8.200e-30 320-351 BL00411G 21.39 6.100e-28 270-312 BL00411C 15.04 7.000e-22 79-101 BL00411F 14.77 1.273e-19 208-233 BL00411E 10.43 7.429e-12 142-161 BL00411A 11.31 4.484e-11 9-24 BL00411B 13.51 1.563e-10 45-62
971	PD00301	PROTEIN REPEAT MUSCLE CALCIUM-BL	PD00301A 10.24 6.400e-09 598-609
971	DM01399	VARICELLA-ZOSTER VIRUS GENE 54 PROTEIN.	DM01399B 12.42 7.092e-09 1571-1583

*Results include in order: accession number subtype; raw score; p-value; position of signature in amino acid sequence.

SEQ ID NO:	Database entry ID	Description	Results*
971	BL00678	Trp-Asp (WD) repeat proteins proteins.	BL00678 9.67 8.800e-10 1350-1361 BL00678 9.67 7.158e-09 1629-1640
971	BL00502	Polygalacturonase proteins.	BL00502A 13.44 7.341e-09 1424-1439
971	BL00412	Neuromodulin (GAP-43) proteins.	BL00412D 16.54 7.796e-09 568-619
971	BL00226	Intermediate filaments proteins.	BL00226B 23.86 8.012e-09 930-978
971	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320B 12.19 2.385e-12 1348-1363 PR00320C 13.01 1.720e-10 1348-1363 PR00320A 16.74 4.971e-10 1348-1363 PR00320B 12.19 5.886e-10 1544-1559 PR00320A 16.74 3.415e-09 1544-1559 PR00320C 13.01 5.500e-09 1498-1513 PR00320B 12.19 8.650e-09 1627-1642 PR00320C 13.01 9.100e-09 1627-1642

SEQ ID NO:	Database entry ID	Description	Results ^a
973	BL01242	Formamidopyrimidine-DNA glycosylase proteins.	BL01242F 17.92 5.300e-11 32-66
975	PF00651	BTB (also known as BR-C/Ttk) domain proteins.	PF00651I 15.00 2.500e-14 46-59
975	PR00501	KELCH REPEAT SIGNATURE	PR00501A 8.25 7.750e-11 371-385
976	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 3.898e-09 99-132
977	PR00380	KINESIN HEAVY CHAIN SIGNATURE	PR00380A 14.18 9.250e-25 93-115 PR00380D 9.93 4.837e-19 302-324 PR00380B 12.64 4.429e-18 212-230 PR00380C 13.18 1.692e-16 247-266
977	BL00411	Kinesin motor domain proteins.	BL00411G 21.39 7.750e-32 248-290 BL00411F 14.77 1.000e-25 203-228 BL00411C 15.04 1.621e-24 93-115 BL00411H 15.66 1.871e-24 301-332 BL00411E 10.43 6.625e-20 143-162 BL00411A 11.31 4.484e-11 5-20
977	PF00846	Hantavirus nucleocapsid protein.	PF00846H 3.96 9.182e-10 408-445
977	BL00224	Clastrin light chain proteins.	BL00224B 16.94 7.136e-09 363-416
978	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 3.368e-18 36-67
978	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 2.068e-10 35-55
980	BL00594	Aromatic amino acids permeases proteins.	BL00594A 16.75 9.376e-09 76-120
982	BL00790	Receptor tyrosine kinase class V proteins.	BL00790E 29.58 1.111e-12 614-662 BL00790F 29.58 3.111e-12 668-716 BL00790E 29.58 7.000e-10 560-608
982	BL00279	Membrane attack complex components / perforin proteins.	BL00279E 37.11 7.632e-12 727-775 BL00279E 37.11 2.047e-09 765-813
982	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 4.600e-11 791-807 BL01187B 12.04 3.571e-10 829-845 BL01187B 12.04 8.714e-10 753-769 BL01187A 9.98 4.375e-09 812-824 BL01187A 9.98 5.125e-09 774-786
982	PR00343	SELECTIN SUPERFAMILY COMPLEMENT-BINDING REPEAT SIGNATURE	PR00343C 16.85 5.364e-09 13-32
982	PR00764	COMPLEMENT C9 SIGNATURE	PR00764F 16.89 8.027e-10 744-765 PR00764F 16.89 6.844e-09 782-803
982	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010A 11.79 6.192e-11 814-826 PR00010C 11.16 5.909e-10 758-769 PR00010A 11.79 7.677e-10 776-788 PR00010C 11.16 1.214e-09 796-807 PR00010C 11.16 8.286e-09 834-845
986	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 4.150e-12 126-149
987	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 6.087e-10 93-107 PR00019B 11.36 7.840e-09 90-104

SEQ ID NO:	Database entry ID	Description	Results*
990	PD01443	INHIBITOR CALPAIN CALPASTATIN REPEAT THIOI PROT.	PD01443D 8.36 4.670e-09 815-837
991	BL01101	Caseln kinase II regulatory subunit proteins.	BL01101A 16.07 1.000e-40 9-54 BL01101B 10.94 9.000e-31 72-97
991	PR00472	CASEIN KINASE II REGULATORY SUBUNIT FAMILY SIGNATURE	PR00472C 12.38 5.154e-28 80-102 PR00472A 8.03 7.600e-23 8-25 PR00472B 14.84 1.000e-19 25-40
994	BL01166	RNA polymerases beta chain proteins.	BL01166G 18.10 2.500e-34 824-866 BL01166H 19.03 9.410e-30 936-986 BL01166D 17.37 4.396e-19 612-642 BL01166E 13.47 8.244e-17 682-706 BL01166C 12.21 9.357e-12 431-456
995	BL00518	Zinc finger, C3HC4 type (RING finger), proteins.	BL00518 12.23 7.000e-09 25-34
996	BL00038	Myc-type, 'helix-loop-helix' dimerization domain proteins.	BL00038A 13.61 6.625e-11 284-300
996	DM00984	w MYOD MYOBLAST DETERMINATION SHORT.	DM00984B 15.18 3.901e-09 262-317
996	BL01009	Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins.	BL01009A 13.75 7.750e-09 260-278
997	BL00905	GTP1/OBG family proteins.	BL00905D 15.00 4.214e-10 125-140
997	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449C 17.27 3.903e-13 63-86 PR00449A 13.20 7.750e-10 22-44
997	BL01019	ADP-ribosylation factors family proteins.	BL01019A 13.20 8.624e-10 56-96
997	BL01115	GTP-binding nuclear protein ran proteins.	BL01115B 10.81 1.505e-09 102-146
998	BL00107	Protein kinases ATP-binding region proteins.	BL00107B 13.31 7.300e-15 64-80
998	BL00239	Receptor tyrosine kinase class II proteins.	BL00239E 17.14 2.693e-10 36-86
1001	BL00383	Tyrosine specific protein phosphatases proteins.	BL00383E 10.35 5.263e-09 328-339
1003	BL00383	Tyrosine specific protein phosphatases proteins.	BL00383E 10.35 5.263e-09 246-257
1004	BL00383	Tyrosine specific protein phosphatases proteins.	BL00383E 10.35 5.263e-09 328-339
1010	BL00226	Intermediate filaments proteins.	BL00226B 23.86 5.919e-09 560-608
1012	PR00322	G10 PROTEIN SIGNATURE	PR00322E 6.62 1.720e-10 30-40
1012	BL00997	G10 protein.	BL00997C 6.36 3.308e-09 29-39
1013	BL00269	Mammalian defensins proteins.	BL00269C 16.52 6.786e-26 110-139 BL00269A 8.53 2.607e-20 45-65 BL00269B 19.17 5.300e-17 72-101
1014	BL01160	Kinesin light chain repeat proteins.	BL01160B 19.54 8.297e-10 6-60
1015	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 5.846e-11 476-493 BL00028 16.07 6.192e-11 989-1005
1015	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 6.087e-10 473-487 PR00048A 10.52 1.000e-09 985-1000
1015	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 6.571e-12 464-477 PD00066 13.92 7.000e-12 977-990 PD00066 13.92 4.600e-09 949-962 PD00066 13.92

SEQ ID NO:	Database entry ID	Description	Results*
			7.300e-09 397-410
1019	BL00289	Pentaxin family proteins.	BL00289E 18.00 4.375e-13 22-37
1019	PR00895	PENTAXIN SIGNATURE	PR00895G 14.55 4.913e-10 19-31
1022	BL00348	p53 tumor antigen proteins.	BL00348F 23.19 4.571e-09 140-183
1023	BL00455	Putative AMP-binding domain proteins.	BL00455 13.31 6.684e-13 248-264
1023	PR00154	AMP-BINDING SIGNATURE	PR00154A 8.88 7.375e-10 241-253
1026	BL00421	Transmembrane 4 family proteins.	BL00421E 20.97 1.851e-09 17-47
1026	PR00259	TRANSMEMBRANE FOUR FAMILY SIGNATURE	PR00259D 13.50 7.097e-09 20-47
1028	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 2.149e-29 6-45
1028	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 1.692e-15 234-247 PD00066 13.92 3.400e-14 150-163 PD00066 13.92 5.800e-14 206-219 PD00066 13.92 8.714e-12 178-191
1028	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 2.350e-13 102-119 BL00028 16.07 5.500e-11 190-207 BL00028 16.07 6.192e-11 218-235 BL00028 16.07 1.000e-09 134-151
1028	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 4.750e-14 215-229 PR00048A 10.52 5.500e-13 99-113 PR00048A 10.52 4.316e-11 131-145 PR00048B 6.02 4.462e-11 115-125 PR00048A 10.52 1.391e-10 159-173 PR00048B 6.02 6.625e-10 203-213 PR00048A 10.52 4.600e-09 187-201 PR00048B 6.02 8.579e-09 147-157
1029	PR00014	FIBRONECTIN TYPE III REPEAT SIGNATURE	PR00014D 12.04 2.059e-10 215-230
1029	BL00790	Receptor tyrosine kinase class V proteins.	BL00790F 15.90 2.519e-09 157-184
1032	BL00269	Mammalian defensins proteins.	BL00269C 16.52 6.786e-26 133-162 BL00269A 8.53 2.607e-20 68-88 BL00269B 19.17 5.500e-17 95-124
1033	PD01876	ANTIGEN MELANOMA-ASSOCIATED MULTIGENE FAMILY TUM.	PD01876C 21.73 1.231e-20 75-128
1034	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 8.835e-11 326-359
1034	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 4.150e-11 112-126 PR00019B 11.36 8.000e-11 109-123 PR00019B 11.36 8.500e-11 184-198 PR00019A 11.19 6.478e-10 187-201 PR00019A 11.19 7.333e-09 63-77
1034	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 9.609e-09 334-344
1035	PR00988	URIDINE KINASE SIGNATURE	PR00988A 6.39 6.276e-12 338-376
1035	BL00567	Phosphoribulokinase proteins.	BL00567A 10.66 6.459e-11 360-

SEQ ID NO:	Database entry ID	Description	Results*
			379
1038	BL00120	Lipases, serine proteins.	BL00120C 12.62 9.053e-12 95-106
1038	PR00825	VESPID VENOM ALLERGEN PHOSPHOLIPASE A1 SIGNATURE	PR00825B 14.81 1.371e-10 83-104
1038	PR00821	TRIACYLGLYCEROL LIPASE FAMILY SIGNATURE	PR00821E 9.64 4.896e-12 19-38 PR00821F 16.10 1.806e-09 118-134
1044	BL00027	'Homeobox' domain proteins.	BL00027 26.43 5.500e-30 169-212
1044	BL00032	'Homeobox' antennapedia-type protein.	BL00032C 11.28 5.179e-16 197-215 BL00032B 10.83 3.060e-15 158-197
1044	PR00024	HOMEODOMAIN SIGNATURE	PR00024C 7.49 8.071e-13 201-211 PR00024B 11.27 7.000e-12 191-202
1044	PR00026	ENGRAILED HOMEODOMAIN SIGNATURE	PR00026A 7.47 9.710e-12 153-171
1044	PR00031	LAMBDA AND OTHER REPRESSOR HELIX-TURN-HELIX SIGNATURE	PR00031B 16.29 4.724e-10 192-209
1044	PR00025	HOMEOTIC ANTENNAPEA PROTEIN SIGNATURE	PR00025B 11.94 6.434e-10 155-171
1048	BL01128	Shikimate kinase proteins.	BL01128A 18.84 8.200e-14 7-41
1048	BL00211	ABC transporters family proteins.	BL00211A 12.23 4.600e-09 9-21
1049	PR00014	FIBRONECTIN TYPE III REPEAT SIGNATURE	PR00014C 15.44 1.783e-09 211-230 PR00014A 8.22 3.045e-09 373-383 PR00014C 15.44 6.087e-09 309-328
1049	BL00790	Receptor tyrosine kinase class V proteins.	BL00790I 20.01 1.750e-12 642-673 BL00790I 20.01 6.125e-12 320-351 BL00790I 20.01 6.679e-09 222-253
1049	PR00096	GLUTAMINE AMIDOTRANSFERASE SUPERFAMILY SIGNATURE	PR00096B 9.72 9.827e-09 689-701
1050	DM00372	CARCINOEMBRYONIC ANTIGEN PRECURSOR AMINO-TERMINAL DOMAIN.	DM00372C 23.69 4.919e-12 67-103
1053	BL01282	BIR repeat proteins.	BL01282B 30.49 1.000e-11 194-233
1053	PR00544	PROGESTERONE RECEPTOR SIGNATURE	PR00544F 7.76 3.329e-09 18-35
1057	PR00464	E-CLASS P450 GROUP II SIGNATURE	PR00464A 20.47 9.591e-16 149-170 PR00464C 18.84 1.000e-15 324-353 PR00464D 17.40 6.250e-15 353-371 PR00464B 20.41 1.844e-12 205-224
1057	PR00385	P450 SUPERFAMILY SIGNATURE	PR00385A 14.97 1.346e-12 335-353 PR00385B 10.22 4.130e-11 353-367
1057	PR00463	E-CLASS P450 GROUP I SIGNATURE	PR00463E 17.37 4.814e-11 344-371
1058	PD02382	RECEPTOR CHAIN PRECURSOR TRANSME.	PD02382A 17.43 9.321e-09 99-115
1060	BL00795	Involucrin proteins.	BL00795C 17.06 6.442e-10 905-950
1060	BL00422	Granins proteins.	BL00422C 16.18 4.255e-10 910-938 BL00422C 16.18 4.353e-09

SEQ ID NO:	Database entry ID	Description	Results*
			913-941
1060	PF00992	Troponin.	PF00992A 16.67 2.184e-09 900-935 PF00992A 16.67 5.382e-09 889-924
1060	BL00412	Neuromodulin (GAP-43) proteins.	BL00412D 16.54 7.429e-09 883-934
1060	PF01140	Matrix protein (MA), p15.	PF01140D 15.54 8.326e-09 903-938
1063	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 6.644e-09 603-636
1063	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 9.643e-10 96-111 PR00049D 0.00 2.525e-09 95-110 PR00049D 0.00 3.898e-09 114-129 PR00049D 0.00 7.407e-09 97-112
1063	PR00239	MOLLUSCAN RHODOPSIN C-TERMINAL TAIL SIGNATURE	PR00239E 1.58 8.670e-09 100-112
1064	BL01282	BiR repeat proteins.	BL01282B 30.49 1.000e-11 137-176
1064	PR00544	PROGESTERONE RECEPTOR SIGNATURE	PR00544F 7.76 3.329e-09 18-35
1065	BL01282	BiR repeat proteins.	BL01282B 30.49 1.000e-11 187-226
1065	PR00544	PROGESTERONE RECEPTOR SIGNATURE	PR00544F 7.76 3.329e-09 18-35
1066	BL00218	Amino acid permeases proteins.	BL00218D 21.49 7.324e-11 226-271 BL00218E 23.30 3.475e-09 307-347
1067	BL00994	Bacterial export FHIPeP family proteins.	BL00994A 15.15 1.086e-09 71-118
1068	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 8.385e-33 6-45
1068	BL01030	RNA polymerases M / I5 Kd subunits proteins.	BL01030 23.44 7.480e-10 283-321
1068	BL00466	TFIIIS zinc ribbon domain proteins.	BL00466 25.88 5.622e-09 283-320
1068	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 2.385e-15 271-284 PD00066 13.92 3.077e-15 243-256 PD00066 13.92 3.077e-15 320-333 PD00066 13.92 3.077e-15 348-361 PD00066 13.92 3.077e-15 376-389 PD00066 13.92 3.077e-15 432-445 PD00066 13.92 3.077e-15 617-630 PD00066 13.92 3.077e-15 701-714 PD00066 13.92 7.923e-15 215-228 PD00066 13.92 8.200e-14 589-602 PD00066 13.92 8.800e-14 729-742 PD00066 13.92 5.714e-12 542-555 PD00066 13.92 9.571e-12 561-574 PD00066 13.92 3.739e-11 404-417 PD00066 13.92 1.692e-10 299-312 PD00066 13.92 2.038e-10 488-501 PD00066 13.92 8.615e-10 645-658 PD00066 13.92 6.700e-09 757-770

SEQ ID NO:	Database entry ID	Description	Results*
1068	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 5.950e-13 388-405 BL00028 16.07 6.400e-13 444-461 BL00028 16.07 1.692e-11 227-244 BL00028 16.07 3.423e-11 332-349 BL00028 16.07 6.538e-11 685-702 BL00028 16.07 7.231e-11 713-730 BL00028 16.07 7.577e-11 573-590 BL00028 16.07 4.300e-10 601-618 BL00028 16.07 5.500e-10 171-188 BL00028 16.07 5.800e-10 255-272 BL00028 16.07 7.900e-10 657-674 BL00028 16.07 9.700e-10 526-543 BL00028 16.07 2.029e-09 283-300 BL00028 16.07 3.829e-09 741-758 BL00028 16.07 6.914e-09 360-377 BL00028 16.07 7.686e-09 416-433
1068	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 4.750e-14 570-584 PR00048A 10.52 2.929e-13 329-343 PR00048A 10.52 5.500e-13 224-238 PR00048A 10.52 1.000e-12 598-612 PR00048A 10.52 5.235e-12 252-266 PR00048A 10.52 5.765e-12 441-455 PR00048A 10.52 7.353e-12 385-399 PR00048A 10.52 7.353e-12 413-427 PR00048A 10.52 7.353e-12 710-724 PR00048A 10.52 3.842e-11 357-371 PR00048A 10.52 5.263e-11 626-640 PR00048A 10.52 5.737e-11 280-294 PR00048A 10.52 7.632e-11 682-696 PR00048A 10.52 8.579e-11 654-668 PR00048B 6.02 2.125e-10 457-467 PR00048A 10.52 2.565e-10 738-752 PR00048A 10.52 6.087e-10 523-537 PR00048B 6.02 7.188e-10 698-708 PR00048B 6.02 7.750e-10 726-736 PR00048B 6.02 1.000e-09 240-250 PR00048A 10.52 3.520e-09 308-322 PR00048A 10.52 4.600e-09 551-565 PR00048A 10.52 8.560e-09 196-210 PR00048B 6.02 1.000e-08 586-596
1069	PD01427	TRANSFERASE METHYLTRANSFERASE BL	PD01427B 22.45 1.545e-11 117-158
1070	PF00168	C2 domain proteins.	PF00168C 27.49 1.750e-09 202-228
1070	PR00360	C2 DOMAIN SIGNATURE	PR00360B 13.61 2.227e-09 219-233
1075	PR00962	LETHAL(2) GIANT LARVAE PROTEIN SIGNATURE	PR00962D 10.40 3.054e-10 178-202
1075	PR00320	G-PROTEIN BETA WD-40 REPEAT	PR00320A 16.74 2.976e-09 181-

SEQ ID NO:	Database entry ID	Description	Results*
		SIGNATURE	196
1078	DM01970	0 kw ZK632.12 YDR313C ENDOSOMAL III.	DM01970B 8.60 2.478e-13 310-323
1078	BL00678	Trp-Asp (WD) repeat proteins proteins.	BL00678 9.67 1.474e-09 41-52
1078	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320A 16.74 1.931e-11 214-229 PR00320B 12.19 3.829e-10 214-229 PR00320C 13.01 3.880e-10 214-229 PR00320C 13.01 4.900e-09 257-272
1079	PF00774	Dihydropyridine sensitive L-type calcium channel (Beta subunit).	PF00774D 10.59 8.396e-09 339-365
1079	BL01013	Oxysterol-binding protein family proteins.	BL01013D 26.81 8.839e-09 588-632
1080	BL00615	C-type lectin domain proteins.	BL00615A 16.68 3.160e-11 129-147
1081	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 2.705e-11 47-86
1082	BL00359	Ribosomal protein L11 proteins.	BL00359B 23.07 7.462e-24 160-201 BL00359C 22.18 6.586e-22 215-249 BL00359A 20.66 4.000e-21 124-160
1082	BL01108	Ribosomal protein L24 proteins.	BL01108A 20.33 1.000e-08 144-177
1084	PD02462	PROTEIN BOLA TRANSCRIPTION REGULATION AC.	PD02462A 22.48 1.220e-09 104-139
1084	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 7.000e-17 755-772 BL00028 16.07 6.625e-15 699-716 BL00028 16.07 8.412e-14 223-240 BL00028 16.07 8.941e-14 167-184 BL00028 16.07 6.850e-13 391-408 BL00028 16.07 1.783e-12 559-576 BL00028 16.07 2.957e-12 307-324 BL00028 16.07 7.652e-12 503-520 BL00028 16.07 7.652e-12 811-828 BL00028 16.07 8.043e-12 335-352 BL00028 16.07 1.346e-11 447-464 BL00028 16.07 2.385e-11 867-884 BL00028 16.07 4.462e-11 671-688 BL00028 16.07 5.846e-11 587-604 BL00028 16.07 6.192e-11 839-856 BL00028 16.07 6.192e-11 895-912 BL00028 16.07 8.962e-11 531-548 BL00028 16.07 1.600e-10 279-296 BL00028 16.07 1.600e-10 363-380 BL00028 16.07 6.100e-10 111-128 BL00028 16.07 6.700e-10 643-660 BL00028 16.07 8.500e-10 251-268 BL00028 16.07 1.771e-09 783-800 BL00028 16.07 5.886e-09 475-492
1084	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 4.600e-18 696-710 PR00048A 10.52 5.091e-15 164-178 PR00048A 10.52 6.727e-15 836-850 PR00048A 10.52

SEQ ID NO:	Database entry ID	Description	Results*
			1.000e-14 220-234 PR00048A 10.52 2.500e-14 444-458 PR00048A 10.52 5.500e-14 556-570 PR00048A 10.52 9.250e-14 388-402 PR00048A 10.52 1.643e-13 640-654 PR00048A 10.52 3.571e-13 304-318 PR00048A 10.52 3.571e-13 528-542 PR00048A 10.52 8.071e-13 332-346 PR00048A 10.52 8.071e-13 808-822 PR00048A 10.52 8.071e-13 864-878 PR00048A 10.52 8.714e-13 500-514 PR00048A 10.52 7.353e-12 892-906 PR00048B 6.02 1.000e-11 292-302 PR00048B 6.02 1.000e-11 516-526 PR00048B 6.02 1.000e-11 824-834 PR00048A 10.52 3.842e-11 276-290 PR00048A 10.52 6.684e-11 584-598 PR00048A 10.52 9.053e-11 668-682 PR00048A 10.52 4.130e-10 360-374 PR00048A 10.52 6.870e-10 752-766 PR00048B 6.02 1.474e-09 768-778 PR00048B 6.02 3.368e-09 236-246 PR00048B 6.02 3.368e-09 460-470 PR00048B 6.02 4.789e-09 152-162 PR00048B 6.02 4.789e-09 376-386 PR00048B 6.02 4.789e-09 600-610 PR00048A 10.52 4.960e-09 108-122 PR00048A 10.52 4.960e-09 248-262 PR00048B 6.02 6.211e-09 264-274 PR00048B 6.02 6.211e-09 488-498 PR00048B 6.02 6.211e-09 796-806 PR00048B 6.02 7.632e-09 404-414 PR00048B 6.02 7.632e-09 712-722 PR00048A 10.52 8.920e-09 780-794
1084	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 1.600e-14 295-308 PD00066 13.92 1.600e-14 323-336 PD00066 13.92 1.600e-14 519-532 PD00066 13.92 1.600e-14 547-560 PD00066 13.92 1.600e-14 827-840 PD00066 13.92 1.600e-14 855-868 PD00066 13.92 5.200e-14 351-364 PD00066 13.92 5.200e-14 575-588 PD00066 13.92 8.200e-14 883-896 PD00066 13.92 9.400e-14 239-252 PD00066 13.92 2.500e-13 155-168 PD00066 13.92 2.500e-13 379-392 PD00066 13.92 2.286e-12 267-280 PD00066 13.92 2.286e-12 491-504 PD00066

SEQ ID NO:	Database entry ID	Description	Results*
			13.92 2.286e-12 799-812 PD00066 13.92 2.714e-12 463-476 PD00066 13.92 2.714e-12 715-728 PD00066 13.92 2.714e-12 771-784 PD00066 13.92 3.571e-12 687-700 PD00066 13.92 7.000e-12 407-420 PD00066 13.92 1.000e-10 127-140 PD00066 13.92 1.000e-08 603-616
1085	PR00679	PROHIBITIN SIGNATURE	PR00679F 8.03 6.478e-28 178-202 PR00679C 14.44 7.677e-22 107-126 PR00679E 12.82 5.171e-19 153-173 PR00679D 11.91 9.053e-18 130-147 PR00679G 6.13 7.882e-17 201-218 PR00679B 13.63 2.444e-10 84-104
1086	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245E 12.40 8.286e-12 45-60
1086	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 5.814e-09 26-53
1087	PD00126	PROTEIN REPEAT DOMAIN TPR NUCLEA.	PD00126A 22.53 6.885e-10 99-120
1088	BL01145	Ribosomal protein L34e proteins.	BL01145A 13.73 1.000e-12 3-45
1093	BL00154	E1-E2 ATPases phosphorylation site proteins.	BL00154D 12.57 7.387e-09 95-106
1093	PR00121	SODIUM/POTASSIUM-TRANSPORTING ATPASE SIGNATURE	PR00121E 13.97 9.444e-09 92-111
1095	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 7.923e-15 439-452 PD00066 13.92 2.800e-14 411-424 PD00066 13.92 2.800e-14 467-480 PD00066 13.92 5.800e-14 495-508 PD00066 13.92 5.800e-14 523-536 PD00066 13.92 8.200e-14 355-368 PD00066 13.92 5.500e-13 579-592 PD00066 13.92 3.143e-12 551-564 PD00066 13.92 4.857e-12 383-396
1095	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 4.150e-13 367-384 BL00028 16.07 8.200e-13 563-580 BL00028 16.07 3.348e-12 479-496 BL00028 16.07 7.652e-12 423-440 BL00028 16.07 8.826e-12 619-636 BL00028 16.07 4.115e-11 451-468 BL00028 16.07 5.500e-11 395-412 BL00028 16.07 7.231e-11 591-608 BL00028 16.07 1.600e-10 339-356 BL00028 16.07 2.200e-10 335-552
1095	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 4.375e-16 560-574 PR00048A 10.52 4.214e-13 476-490 PR00048A 10.52 6.143e-13 364-378 PR00048B 6.02 6.400e-13 492-502 PR00048B 6.02 1.000e-11 352-362 PR00048B 6.02 1.000e-11 408-

SEQ ID NO:	Database entry ID	Description	Results*
			418 PR00048B 6.02 1.692e-11 548-558 PR00048A 10.52 1.947e-11 588-602 PR00048A 10.52 3.842e-11 420-434 PR00048B 6.02 4.462e-11 464-474 PR00048A 10.52 6.684e-11 392-406 PR00048A 10.52 6.684e-11 448-462 PR00048B 6.02 7.231e-11 436-446 PR00048A 10.52 5.696e-10 532-546 PR00048B 6.02 5.263e-09 576-586 PR00048A 10.52 6.400e-09 504-518 PR00048A 10.52 6.760e-09 336-350 PR00048A 10.52 7.120e-09 616-630
1095	PD02462	PROTEIN BOLA TRANSCRIPTION REGULATION AC.	PD02462A 22.48 9.232e-09 472-507
1097	BL00649	G-protein coupled receptors family 2 proteins.	BL00649C 17.82 9.542e-12 400-426
1097	PR00249	SECRETIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00249C 17.08 8.839e-11 402-426 PR00249A 15.88 7.851e-09 330-355
1097	PR00248	METABOTROPIC GLUTAMATE GPCR SIGNATURE	PR00248E 17.85 9.366e-09 442-465
1100	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479B 12.57 8.875e-09 886-902
1101	PR00205	CADHERIN SIGNATURE	PR00205B 11.39 5.655e-16 255-273 PR00205A 14.73 1.000e-12 180-196 PR00205B 11.39 4.927e-10 475-493 PR00205C 13.65 9.438e-10 515-530
1101	BL00232	Cadherins extracellular repeat proteins domain proteins.	BL00232B 32.79 1.000e-40 148-196 BL00232A 27.72 5.125e-25 54-87 BL00232B 32.79 4.286e-19 257-305 BL00232C 10.65 7.429e-16 255-273 BL00232B 32.79 1.500e-10 372-420 BL00232C 10.65 6.538e-10 475-493 BL00232C 10.65 7.632e-09 146-164
1103	BL00122	Carboxylesterases type-B serine proteins.	BL00122A 12.04 3.152e-15 86-107 BL00122D 12.53 7.097e-14 197-213 BL00122B 16.84 1.346e-13 148-159 BL00122C 7.91 9.550e-10 168-179
1105	BL00018	EF-hand calcium-binding domain proteins.	BL00018 7.41 2.800e-10 217-230 BL00018 7.41 8.650e-10 133-146
1105	PR00450	RECOVERIN FAMILY SIGNATURE	PR00450E 12.14 3.438e-15 174-193 PR00450B 11.76 5.574e-13 82-102 PR00450D 16.58 6.714e-13 152-172 PR00450C 12.22 6.864e-12 128-150 PR00450G 15.33 6.591e-09 224-245
1108	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 6.143e-12 927-940 PD00066 13.92 7.000e-09 343-356
1108	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 3.250e-13 911-928 BL00028 16.07 9.100e-13

SEQ ID NO:	Database entry ID	Description	Results*
			327-344 BL00028 16.07 3.348e-12 939-956 BL00028 16.07 5.500e-11 270-287 BL00028 16.07 4.000e-10 298-315 BL00028 16.07 6.700e-10 968-985 BL00028 16.07 9.700e-10 191-208 BL00028 16.07 9.700e-10 355-372 BL00028 16.07 4.857e-09 384-401 BL00028 16.07 7.171e-09 242-259
1108	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048B 6.02 6.000e-12 924-934 PR00048A 10.52 4.522e-10 324-338 PR00048A 10.52 4.913e-10 908-922 PR00048A 10.52 7.652e-10 936-950 PR00048A 10.52 4.600e-09 965-979 PR00048A 10.52 7.480e-09 352-366
1108	DM01970	0 kw ZK632.12 YDR313C ENDOSOMAL III.	DM01970A 8.50 9.100e-09 910-917
1109	DM01857	5 kw NUCLEOSIDE TRANSPORT DEPENDENT NA.	DM01857B 14.94 6.471e-19 284-312 DM01857E 18.90 7.313e-18 488-527 DM01857F 12.86 7.045e-15 548-575 DM01857C 15.62 4.500e-14 312-344 DM01857A 20.25 1.667e-13 207-250 DM01857D 16.80 3.165e-12 372-410
1112	DM01840	kw SPAC24B11.09 R07E5.13.	DM01840B 22.04 1.844e-40 59-103 DM01840A 10.95 9.571e-13 31-43
1114	BL01115	GTP-binding nuclear protein ran proteins.	BL01115A 10.22 3.438e-14 53-97
1114	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 9.308e-16 53-75 PR00449C 17.27 8.920e-15 94-117 PR00449B 14.34 5.680e-10 76-93
1114	PR00879	FISH ACETYLCHOLINESTERASE SIGNATURE	PR00879A 6.28 1.450e-09 37-43
1114	BL01125	ROK family proteins.	BL01125D 13.61 7.429e-09 214-227
1115	PF00622	Domain in SPlA and the RYanodine Receptor.	PF00622B 21.00 2.500e-13 265-287
1115	BL00518	Zinc finger, C3HC4 type (RING finger), proteins.	BL00518 12.23 9.571e-10 103-112
1116	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 5.050e-11 66-80 PR00019B 11.36 6.850e-10 63-77
1116	DM00315	072 RIBONUCLEASE INHIBITOR.	DM00315G 15.85 6.362e-10 84-120 DM00315G 15.85 3.340e-09 246-282
1117	BL00509	Ras GTPase-activating proteins.	BL00509B 10.28 5.263e-10 429-440
1117	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 9.357e-10 794-809 PR00049D 0.00 1.915e-09 793-808 PR00049D 0.00 3.593e-09 792-807 PR00049D 0.00 5.729e-09 791-806
1117	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 7.559e-09 780-813

SEQ ID NO:	Database entry ID	Description	Results*
1117	PR00806	VINCULIN SIGNATURE	PR00806A 6.63 8.397e-09 794-805
1120	PR00720	MAMMALIAN LMW PHOSPHOTYROSINE PROTEIN PHOSPHATASE SIGNATURE	PR00720C 12.41 1.099e-27 88-109 PR00720B 10.61 4.789e-20 71-87 PR00720A 16.54 2.000e-17 28-41 PR00720E 10.01 1.342e-16 117-139 PR00720D 17.32 1.778e-15 110-127
1120	PR00719	LMW PHOSPHOTYROSINE PROTEIN PHOSPHATASE SIGNATURE	PR00719A 14.49 3.000e-23 9-27 PR00719C 14.10 5.000e-18 85-101 PR00719B 14.32 1.346e-15 52-69 PR00719D 17.52 9.654e-15 110-124
1121	PD00131	ATP-BINDING TRANSPORT TRANSMEMBR.	PD00131B 34.97 7.987e-09 108-162
1123	BL00615	C-type lectin domain proteins.	BL00615A 16.68 9.526e-13 573-591
1123	PR00356	TYPE II ANTIFREEZE PROTEIN SIGNATURE	PR00356C 13.33 3.793e-10 591-609 PR00356D 13.09 5.038e-09 619-636
1123	PR00439	11-S SEED STORAGE PROTEIN FAMILY SIGNATURE	PR00439C 15.32 9.217e-09 332-353
1127	BL00134	Serine proteases, trypsin family, histidine proteins.	BL00134A 11.96 5.781e-15 493-510 BL00134B 15.99 4.194e-14 675-699
1127	BL00021	Kringle domain proteins.	BL00021B 13.33 8.984e-12 493-511
1127	BL00495	Apple domain proteins.	BL00495N 11.04 9.735e-11 667-702
1127	PR00453	VON WILLEBRAND FACTOR TYPE A DOMAIN SIGNATURE	PR00453A 12.79 5.345e-14 251-269 PR00453B 14.65 4.682e-10 290-305
1127	PR00722	CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1) SIGNATURE	PR00722A 12.27 9.129e-13 494-510 PR00722C 10.87 4.273e-11 674-687 PR00722B 12.51 4.000e-09 554-569
1127	BL01253	Type 1 fibronectin domain proteins.	BL01253G 11.34 5.348e-09 674-688
1128	BL00236	Neurotransmitter-gated ion-channels proteins.	BL00236D 25.66 4.000e-30 64-106
1128	PR00252	NEUROTRANSMITTER-GATED ION CHANNEL FAMILY SIGNATURE	PR00252D 12.29 7.097e-10 71-84
1129	BL00604	Synaptophysin / synaptoporin proteins.	BL00604F 5.96 7.718e-10 367-412
1129	PR00524	CHOLECYSTOKININ TYPE A RECEPTOR SIGNATURE	PR00524F 5.36 7.415e-09 208-222
1129	BL01113	C1q domain proteins.	BL01113A 17.99 6.455e-14 158-185 BL01113A 17.99 5.622e-13 28-55 BL01113A 17.99 1.923e-12 393-420 BL01113A 17.99 4.462e-12 230-257 BL01113A 17.99 6.538e-12 34-61 BL01113A 17.99 1.000e-11 284-311 BL01113A 17.99 1.205e-11 31-58 BL01113A 17.99 1.614e-11 170-197 BL01113A 17.99 5.091e-11 167-194 BL01113A 17.99 6.523e-11 456-483

SEQ ID NO:	Database entry ID	Description	Results*
			BL01113A 17.99 7.341e-11 411-438 BL01113A 17.99 1.766e-10 402-429 BL01113A 17.99 2.915e-10 356-383 BL01113A 17.99 6.745e-10 239-266 BL01113A 17.99 7.319e-10 293-320 BL01113A 17.99 8.085e-10 164-191 BL01113A 17.99 8.660e-10 305-332 BL01113A 17.99 2.038e-09 272-299 BL01113A 17.99 2.212e-09 353-380 BL01113A 17.99 2.385e-09 86-113 BL01113A 17.99 2.731e-09 149-176 BL01113A 17.99 2.904e-09 453-480 BL01113A 17.99 3.423e-09 131-158 BL01113A 17.99 3.423e-09 308-335 BL01113A 17.99 3.769e-09 396-423 BL01113A 17.99 3.769e-09 450-477 BL01113A 17.99 3.942e-09 25-52 BL01113A 17.99 3.942e-09 173-200 BL01113A 17.99 4.115e-09 278-305 BL01113A 17.99 4.981e-09 43-70 BL01113A 17.99 5.327e-09 19-46 BL01113A 17.99 5.500e-09 320-347 BL01113A 17.99 6.192e-09 143-170 BL01113A 17.99 7.231e-09 224-251 BL01113A 17.99 8.269e-09 245-272
1129	BL00420	Speract receptor repeat proteins domain proteins.	BL00420A 20.42 3.571e-13 31-60 BL00420A 20.42 9.082e-13 113-142 BL00420A 20.42 8.691e-11 311-340 BL00420A 20.42 4.098e-10 125-154 BL00420A 20.42 4.541e-10 158-187 BL00420A 20.42 5.279e-10 34-63 BL00420A 20.42 5.426e-10 137-166 BL00420A 20.42 6.754e-10 49-78 BL00420A 20.42 6.902e-10 266-295 BL00420A 20.42 7.492e-10 43-72 BL00420A 20.42 8.082e-10 25-54 BL00420A 20.42 9.852e-10 167-196 BL00420A 20.42 2.800e-09 170-199 BL00420A 20.42 2.938e-09 414-443 BL00420A 20.42 3.492e-09 52-81 BL00420A 20.42 5.015e-09 305-334 BL00420A 20.42 5.569e-09 37-66 BL00420A 20.42 7.923e-09 459-488 BL00420A 20.42 8.477e-09 335-364 BL00420A 20.42 8.754e-09 83-112 BL00420A 20.42 9.446e-09 287-316 BL00420A 20.42 9.862e-09 290-319

SEQ ID NO:	Database entry ID	Description	Results*
1132	PR00042	FOS TRANSFORMING PROTEIN SIGNATURE	PR00042E 9.69 7.652e-09 234-258
1135	PD00066	PROTEIN ZINC-FINGER METAL-BINDI	PD00066 13.92 9.400e-14 365-378 PD00066 13.92 6.143e-12 335-348 PD00066 13.92 2.174e-11 395-408
1135	BL00970	Nuclear transition protein 2 proteins.	BL00970B 10.09 3.069e-10 55-81
1135	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 3.746e-09 223-238 PR00049D 0.00 3.746e-09 224-239 PR00049D 0.00 3.898e-09 220-235
1135	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 6.885e-11 349-366 BL00028 16.07 5.886e-09 379-396
1135	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048B 6.02 6.211e-09 362-372
1135	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 8.169e-09 221-254
1136	BL00962	Ribosomal protein S2 proteins.	BL00962D 22.51 5.500e-35 131-175 BL00962C 15.90 9.591e-17 106-124 BL00962B 36.15 9.060e-15 40-94
1136	PR00395	RIBOSOMAL PROTEIN S2 SIGNATURE	PR00395C 16.17 1.000e-17 106-124 PR00395D 13.04 7.000e-17 131-149 PR00395E 10.56 6.400e-16 169-184 PR00395E 14.46 4.103e-11 148-160
1137	BL00152	ATP synthase alpha and beta subunits proteins.	BL00152A 15.38 5.109e-14 128-154
1139	BL00152	ATP synthase alpha and beta subunits proteins.	BL00152B 21.40 4.273e-37 124-162 BL00152A 15.38 8.364e-23 67-93
1139	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 2.862e-09 139-161
1140	BL00152	ATP synthase alpha and beta subunits proteins.	BL00152B 21.40 2.000e-32 185-223 BL00152A 15.38 8.364e-23 128-154
1140	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 7.672e-09 200-222
1141	PR00493	BREAST CANCER TYPE I SUSCEPTIBILITY PROTEIN SIGNATURE	PR00493G 7.57 1.184e-10 652-673
1141	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 3.700e-10 494-507
1141	BL00518	Zinc finger, C3HC4 type (RING finger), proteins.	BL00518 12.23 3.667e-09 102-111
1141	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 9.047e-15 501-556 PF00791B 28.49 9.386e-14 468-523 PF00791C 20.98 9.814e-10 515-554 PF00791C 20.98 7.618e-09 482-521
1141	PF00023	Ank repeat proteins.	PF00023A 16.03 3.500e-12 534-550 PF00023A 16.03 7.857e-11 501-517 PF00023B 14.20 9.591e-09 497-507
1143	BL00301	GTP-binding elongation factors proteins.	BL00301A 12.41 1.750e-12 72-84
1143	PR00315	GTP-BINDING ELONGATION FACTOR SIGNATURE	PR00315A 11.81 4.000e-14 72-86 PR00315B 11.66 7.600e-10 118-

SEQ ID NO:	Database entry ID	Description	Results*
			127
1145	BL00745	Prokaryotic-type class I peptide chain release factors signal.	BL00745C 13.66 7.398e-18 59-106
1146	BL00745	Prokaryotic-type class I peptide chain release factors signal.	BL00745C 13.66 4.706e-12 59-106
1149	BL00660	Band 4.1 family domain proteins.	BL00660B 17.33 4.800e-27 136-176 BL00660A 31.50 7.911e-20 52-105 BL00660C 23.36 2.241e-19 215-259 BL00660E 23.41 9.647e-13 301-324
1149	PR00935	BAND 4.1 PROTEIN FAMILY SIGNATURE	PR00935C 11.98 4.300e-17 154-175 PR00935D 10.20 1.281e-14 215-232 PR00935B 10.58 6.108e-12 141-155 PR00935A 10.16 3.605e-10 76-89
1149	PR00661	ERM FAMILY SIGNATURE	PR00661C 9.53 3.616e-10 150-172
1153	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 1.882e-12 155-174
1153	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 1.818e-15 155-186 BL00107B 13.31 8.714e-11 221-237
1153	BL00239	Receptor tyrosine kinase class II proteins.	BL00239B 25.15 1.774e-09 90-138
1153	BL00240	Receptor tyrosine kinase class III proteins.	BL00240E 11.56 6.657e-09 141-179
1153	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479C 12.01 9.000e-09 200-213
1155	PR00837	ALLERGEN V5/TPX-1 FAMILY SIGNATURE	PR00837C 17.21 4.064e-11 155-172 PR00837A 14.77 4.960e-10 78-97 PR00837B 11.64 1.310e-09 133-147
1155	BL01009	Extracellular proteins SCP/TPX-1/Ag5/PR-1/Sc7 proteins.	BL01009D 14.19 8.759e-12 156-177 BL01009C 10.54 1.730e-09 133-147
1155	PR00199	ANNEXIN TYPE III SIGNATURE	PR00199F 16.19 9.483e-09 113-140
1156	BL01115	GTP-binding nuclear protein ran proteins.	BL01115A 10.22 8.909e-13 4-48
1156	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 2.059e-19 4-26 PR00449C 17.27 1.000e-18 44-67 PR00449B 14.34 6.727e-11 27-44
1159	BL00175	Phosphoglycerate mutase family phosphohistidine proteins.	BL00175D 27.67 4.000e-40 367-419 BL00175C 23.75 6.870e-28 316-348 BL00175A 15.42 8.200e-19 252-272 BL00175B 12.60 8.714e-17 299-312
1159	BL00300	SRP54-type proteins GTP-binding domain proteins.	BL00300B 20.56 7.534e-11 38-84
1160	PD00066	PROTEIN ZINC-FINGER METAL-BINDI	PD00066 13.92 1.000e-14 547-560 PD00066 13.92 2.200e-14 353-366 PD00066 13.92 3.400e-14 241-254 PD00066 13.92 6.400e-14 325-338 PD00066 13.92 1.500e-13 297-310 PD00066 13.92 6.500e-13 465-478 PD00066 13.92 7.500e-13 437-450 PD00066 13.92 8.500e-13 409-422 PD00066 13.92

SEQ ID NO:	Database entry ID	Description	Results*
			2.714e-12 269-282 PD00066 13.92 3.571e-12 381-394 PD00066 13.92 7.577e-10 519-532
1160	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 3.647e-14 531-548 BL00028 16.07 9.471e-14 309-326 BL00028 16.07 1.900e-13 559-576 BL00028 16.07 7.750e-13 477-494 BL00028 16.07 2.174e-12 337-354 BL00028 16.07 6.478e-12 225-242 BL00028 16.07 8.043e-12 421-438 BL00028 16.07 9.217e-12 365-382 BL00028 16.07 2.038e-11 253-270 BL00028 16.07 7.231e-11 281-298 BL00028 16.07 6.100e-10 449-466 BL00028 16.07 8.800e-10 503-520 BL00028 16.07 5.371e-09 393-410
1160	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 7.429e-13 418-432 PR00048A 10.52 2.588e-12 362-376 PR00048A 10.52 3.647e-12 556-570 PR00048A 10.52 8.412e-12 474-488 PR00048A 10.52 9.471e-12 222-236 PR00048A 10.52 2.421e-11 306-320 PR00048B 6.02 4.462e-11 350-360 PR00048A 10.52 5.737e-11 250-264 PR00048A 10.52 7.632e-11 528-542 PR00048A 10.52 8.579e-11 278-292 PR00048A 10.52 9.053e-11 446-460 PR00048B 6.02 9.308e-11 544-554 PR00048B 6.02 9.308e-11 572-582 PR00048B 6.02 1.000e-10 406-416 PR00048A 10.52 2.565e-10 334-348 PR00048B 6.02 4.938e-10 378-388 PR00048A 10.52 5.304e-10 390-404 PR00048B 6.02 8.313e-10 238-248 PR00048B 6.02 1.947e-09 434-444 PR00048B 6.02 2.421e-09 462-472 PR00048B 6.02 3.368e-09 322-332 PR00048B 6.02 3.842e-09 266-276 PR00048B 6.02 6.211e-09 294-304
1161	PD02331	CYCLIN CELL CYCLE DIVISION PROTE.	PD02331C 13.84 1.913e-11 9-36
1161	BL00048	Protamine P1 proteins.	BL00048 6.39 3.700e-09 165-192 BL00048 6.39 4.938e-09 281-308
1161	DM01206	CORONAVIRUS NUCLEOCAPSID PROTEIN.	DM01206B 10.69 9.328e-11 169-189 DM01206B 10.69 1.247e-10 248-268 DM01206B 10.69 7.781e-10 200-220 DM01206B 10.69 6.582e-09 246-266
1163	BL00239	Receptor tyrosine kinase class II proteins.	BL00239B 25.15 3.915e-15 100-148

SEQ ID NO:	Database entry ID	Description	Results*
1163	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 6.362e-13 165-184
1163	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479C 12.01 1.000e-10 208-221
1163	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 3.250e-26 165-196 BL00107B 13.31 3.647e-09 230-246
1164	BL01013	Oxysterol-binding protein family proteins.	BL01013D 26.81 9.135e-22 501-545 BL01013A 25.14 4.600e-14 220-256 BL01013C 9.97 4.906e-12 330-340 BL01013B 11.33 3.017e-11 287-298
1167	BL00289	Pentaxin family proteins.	BL00289A 30.36 6.850e-26 25-56 BL00289E 18.00 6.684e-14 78-93
1167	PR00895	PENTAXIN SIGNATURE	PR00895A 14.53 1.563e-15 48-63 PR00895G 14.55 5.846e-12 75-87
1167	PR00468	PLANT LIPOXYGENASE SIGNATURE	PR00468I 13.42 9.870e-09 59-74
1168	PR00217	43 KD POSTSYNAPTIC PROTEIN SIGNATURE	PR00217C 10.91 7.527e-09 547-563
1169	PR00756	MEMBRANE ALANYL DIPEPTIDASE (M1) FAMILY SIGNATURE	PR00756D 10.58 1.529e-21 367-383 PR00756B 14.06 5.737e-16 253-269 PR00756A 12.90 1.237e-13 205-221 PR00756E 11.91 4.094e-13 386-399 PR00756C 11.60 6.108e-11 331-342
1169	BL00142	Neutral zinc metalloproteinases, zinc-binding region proteins.	BL00142 8.38 5.500e-10 367-378
1170	DM01688	2 POLY-IG RECEPTOR.	DM01688I 14.97 6.279e-09 75-123
1172	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308C 3.83 2.523e-10 40-50 PR00308B 3.83 8.892e-10 41-51 PR00308E 3.83 8.892e-10 42-52 PR00308B 4.28 6.671e-09 40-52
1172	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 7.115e-09 30-45
1179	BL00300	SRP54-type proteins GTP-binding domain proteins.	BL00300C 25.57 6.000e-09 215-269
1180	BL00514	Fibrinogen beta and gamma chains C-terminal domain proteins.	BL00514C 17.41 9.463e-19 233-270 BL00514B 14.28 7.750e-12 293-310 BL00514D 15.35 9.824e-11 274-287 BL00514G 15.98 4.273e-10 356-386 BL00514H 14.95 6.217e-09 391-416
1181	BL01158	Macrophage migration inhibitory factor family proteins.	BL01158A 21.81 4.130e-30 2-47 BL01158B 17.07 4.316e-29 47-74
1182	BL00456	Sodium:solute symporter family proteins.	BL00456A 22.59 6.250e-40 46-101 BL00456C 24.55 6.586e-40 184-239 BL00456B 18.94 8.125e-25 122-152 BL00456D 6.92 5.500e-10 476-486
1185	PR00830	ENDOPEPTIDASE LA (LON) SERINE PROTEASE (S16) SIGNATURE	PR00830A 8.41 4.780e-14 241-261
1185	PR00918	CALICIVIRUS NON-STRUCTURAL POLYPROTEIN FAMILY SIGNATURE	PR00918A 13.76 1.976e-10 231-252 PR00918A 13.76 2.084e-10 495-516
1185	PR00300	ATP-DEPENDENT CLP PROTEASE ATP-	PR00300A 9.56 5.857e-12 237-

SEQ ID NO:	Database entry ID	Description	Results*
		BINDING SUBUNIT SIGNATURE	256 PR00300A 9.56 5.909e-09 501-520
1185	BL00370	PEP-utilizing enzymes phosphorylation site proteins.	BL00370A 5.71 6.294e-09 79-87
1185	BL00113	Adenylate kinase proteins.	BL00113A 12.74 7.231e-09 502-519
1185	BL00674	AAA-protein family proteins.	BL00674D 23.41 2.286e-30 324-371 BL00674B 4.46 1.205e-17 234-256 BL00674C 22.60 2.059e-17 531-574 BL00674B 4.46 4.886e-17 498-520 BL00674E 15.24 2.800e-15 402-422 BL00674C 22.60 7.600e-09 270-313
1185	BL00870	Chaperonins clpA/B proteins.	BL00870A 11.78 9.534e-09 210-257
1185	BL00675	Sigma-54 interaction domain proteins ATP-binding region A proteins.	BL00675A 24.86 9.775e-09 237-281
1186	PR00165	ANION EXCHANGER SIGNATURE	PR00165I 10.02 8.412e-14 829-849 PR00165A 9.84 6.423e-13 495-518 PR00165B 15.26 9.090e-11 520-541 PR00165F 10.39 6.663e-10 639-658
1186	BL00291	Prion protein.	BL00291A 4.49 9.675e-10 436-471
1186	BL00219	Anion exchangers family proteins.	BL00219B 14.47 2.707e-24 296-340 BL00219C 17.29 5.426e-23 341-380 BL00219K 12.73 9.100e-23 831-873 BL00219M 9.98 9.299e-23 916-962 BL00219H 10.06 5.705e-21 618-666 BL00219I 6.16 4.968e-20 741-795 BL00219A 17.13 7.833e-19 122-154 BL00219E 11.63 2.988e-16 485-525 BL00219F 10.52 8.953e-14 525-549 BL00219G 12.86 8.163e-13 578-617 BL00219L 18.71 8.423e-13 873-912 BL00219N 10.66 6.942e-12 978-1022 BL00219D 15.15 5.286e-11 380-416 BL00219O 14.02 3.377e-09 1023-1063
1186	PD01168	SYNTHETASE LIQASE PROTEIN ALANYL	PD01168L 9.47 7.833e-09 452-467
1186	DM01724	kw ALLERGEN POLLEN CIM1 HOL-L1.	DM01724 8.14 4.296e-10 446-466 DM01724 8.14 6.447e-09 442-462 DM01724 8.14 7.987e-09 438-458
1187	BL00027	'Homeobox' domain proteins.	BL00027 26.43 7.943e-14 65-108
1187	PR00025	HOMEOTIC ANTENNAPEDIA PROTEIN SIGNATURE	PR00025B 11.94 4.000e-11 51-67
1187	BL00032	'Homeobox' antennapedia-type protein.	BL00032B 10.83 7.660e-10 54-93
1187	PR00028	POU DOMAIN SIGNATURE	PR00028D 17.92 2.286e-09 49-70
1189	BL00115	Eukaryotic RNA polymerase II heptapeptide repeat proteins.	BL00115Z 3.12 2.047e-10 469-518
1192	BL01215	Mrp family proteins.	BL01215A 9.75 2.436e-09 466-493
1192	BL00567	Phosphoribulokinase proteins.	BL00567A 10.66 5.757e-09 470-

SEQ ID NO:	Database entry ID	Description	Results*
			489
1192	PR00364	DISEASE RESISTANCE PROTEIN SIGNATURE	PR00364A 8.19 7.341e-09 470-486
1192	BL00113	Adenylate kinase proteins.	BL00113A 12.74 8.062e-09 472-489
1193	PR00303	PREPROTEIN TRANSLOCASE SECY SUBUNIT SIGNATURE	PR00303G 10.45 8.759e-09 88-111
1197	PF00429	ENV polyprotein (coat polyprotein).	PF00429 31.08 8.015e-16 415-465
1198	BL00415	Synapsins proteins.	BL00415N 4.29 7.115e-10 224-268
1198	PR00239	MOLLUSCAN RHODOPSIN C-TERMINAL TAIL SIGNATURE	PR00239E 1.58 1.307e-09 253-265
1198	PR00211	GLUTELIN SIGNATURE	PR00211B 0.86 7.537e-12 245-266 PR00211B 0.86 2.644e-10 251-272 PR00211B 0.86 4.083e-09 233-254 PR00211B 0.86 7.583e-09 239-260
1198	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 5.688e-12 227-260 DM00215 19.43 6.250e-12 225-258 DM00215 19.43 5.235e-11 232-265 DM00215 19.43 5.941e-11 242-275 DM00215 19.43 4.375e-10 236-269 DM00215 19.43 4.857e-10 222-255 DM00215 19.43 5.179e-10 230-263 DM00215 19.43 8.554e-10 237-270 DM00215 19.43 2.068e-09 215-248 DM00215 19.43 3.898e-09 235-268 DM00215 19.43 4.508e-09 240-273 DM00215 19.43 5.376e-09 231-264 DM00215 19.43 6.339e-09 220-253 DM00215 19.43 9.847e-09 218-251
1200	BL00183	Ubiquitin-conjugating enzymes proteins.	BL00183 28.97 4.326e-22 81-129
1202	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 4.214e-16 108-148 BL00237C 13.19 3.323e-11 245-272 BL00237B 5.28 2.227e-09 182-194
1202	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237F 13.57 4.600e-13 250-275 PR00237E 13.03 1.000e-12 174-198 PR00237G 19.63 7.469e-12 288-315 PR00237B 13.50 7.207e-11 75-97 PR00237C 15.69 7.750e-09 122-145
1203	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 4.214e-16 108-148 BL00237C 13.19 3.323e-11 280-307 BL00237B 5.28 2.227e-09 217-229
1203	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237F 13.57 4.600e-13 285-310 PR00237E 13.03 1.000e-12 209-233 PR00237G 19.63 7.469e-12 323-350 PR00237B 13.50 7.207e-11 75-97 PR00237C 15.69 7.750e-09 122-145
1207	PR00259	TRANSMEMBRANE FOUR FAMILY SIGNATURE	PR00259B 14.81 3.769e-21 50-77 PR00259C 16.40 4.000e-20 77-106 PR00259A 9.27 3.600e-16

SEQ ID NO:	Database entry ID	Description	Results*
			12-36
1207	BL00421	Transmembrane 4 family proteins.	BL00421B 17.62 7.261e-36 56-95 BL00421A 11.79 8.313e-16 8-27
1207	PR00164	ABC-2 TYPE TRANSPORT SYSTEM MEMBRANE PROTEIN SIGNATURE	PR00164D 13.90 1.486e-09 9-34
1208	BL00282	Kazal serine protease inhibitors family proteins.	BL00282 16.88 7.207e-14 562-585
1208	BL00216	Sugar transport proteins.	BL00216B 27.64 3.250e-10 267-317
1209	PF00922	Vesiculovirus phosphoprotein.	PF00922A 19.17 7.724e-09 88-122
1214	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 7.143e-10 17-32
1214	PD01351	PROTEIN REPEAT NEUROFILAMENT TRIPL.	PD01351B 13.72 9.518e-10 18-44 PD01351B 13.72 3.758e-09 24-50
1214	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 5.576e-09 5-38
1214	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 7.857e-09 17-30
1215	BL00612	Osteonectin domain proteins.	BL00612E 13.12 3.947e-11 379-424
1215	BL00484	Thyroglobulin type-I repeat proteins proteins.	BL00484C 17.01 3.854e-11 131-146 BL00484B 9.04 1.491e-10 238-252 BL00484C 17.01 8.560e-10 258-273 BL00484B 9.04 3.850e-09 111-125
1216	BL00223	Annexins repeat proteins domain proteins.	BL00223A 15.59 1.000e-33 147-181 BL00223A 15.59 1.435e-16 75-109 BL00223C 24.79 3.928e-15 134-189
1216	PR00196	ANNEXIN FAMILY SIGNATURE	PR00196C 10.36 3.571e-25 151-173 PR00196A 11.16 7.300e-24 84-107 PR00196B 10.68 4.808e-16 124-141 PR00196A 11.16 6.236e-14 156-179 PR00196E 9.19 1.000e-12 155-176 PR00196G 11.72 5.829e-11 199-213 PR00196C 10.36 7.913e-11 79-101
1216	PR00201	ANNEXIN TYPE V SIGNATURE	PR00201D 10.49 8.729e-14 151-173 PR00201A 6.05 1.976e-13 84-107 PR00201G 11.02 3.847e-12 155-182 PR00201A 6.05 8.241e-12 156-179 PR00201H 12.04 4.889e-10 199-213
1216	PR00198	ANNEXIN TYPE II SIGNATURE	PR00198D 7.65 7.787e-21 151-173 PR00198B 8.71 3.880e-17 84-107 PR00198C 14.32 2.688e-11 124-141 PR00198G 8.09 7.033e-10 155-176
1216	PR00200	ANNEXIN TYPE IV SIGNATURE	PR00200E 10.00 5.030e-19 151-173 PR00200G 9.43 5.546e-14 155-182 PR00200B 7.39 4.653e-11 156-179 PR00200B 7.39 4.857e-10 84-107 PR00200H 13.68 9.663e-10 199-213
1216	PR00202	ANNEXIN TYPE VI SIGNATURE	PR00202D 5.58 6.793e-14 151-173 PR00202G 8.01 5.545e-13 155-182 PR00202B 11.44 2.782e-10 155-179 PR00202B 11.44

SEQ ID NO:	Database entry ID	Description	Results*
			5.206e-09 83-107
1216	PR00197	ANNEXIN TYPE I SIGNATURE	PR00197B 7.56 4.960e-29 84-107 PR00197D 7.50 3.000e-26 151-173 PR00197A 8.68 7.577e-20 35-51 PR00197C 7.50 1.000e-19 124-141 PR00197F 9.03 7.128e-10 155-176 PR00197D 7.50 6.250e-09 79-101
1216	PR00199	ANNEXIN TYPE III SIGNATURE	PR00199D 5.65 9.297e-17 151-173 PR00199B 6.86 2.915e-13 84-107 PR00199B 6.86 1.265e-11 156-179 PR00199G 9.09 4.351e-11 156-182 PR00199D 5.65 3.641e-09 79-101 PR00199C 13.84 9.571e-09 124-141
1217	BL00223	Annexins repeat proteins domain proteins.	BL00223B 28.47 1.000e-40 188-238 BL00223A 15.59 1.000e-33 119-153 BL00223A 15.59 1.435e-16 47-81 BL00223C 24.79 3.928e-15 106-161
1217	PR00201	ANNEXIN TYPE V SIGNATURE	PR00201D 10.49 8.729e-14 123-145 PR00201A 6.05 1.976e-13 56-79 PR00201G 11.02 3.847e-12 127-154 PR00201A 6.05 8.241e-12 128-151 PR00201E 12.37 3.317e-11 206-233 PR00201H 12.04 4.889e-10 171-185
1217	PR00198	ANNEXIN TYPE II SIGNATURE	PR00198D 7.65 7.787e-21 123-145 PR00198B 8.71 3.880e-17 56-79 PR00198E 14.67 5.286e-15 206-233 PR00198C 14.32 2.688e-11 96-113 PR00198G 8.09 7.033e-10 127-148
1217	PR00196	ANNEXIN FAMILY SIGNATURE	PR00196D 21.86 1.000e-27 206-233 PR00196C 10.36 3.571e-25 123-145 PR00196A 11.16 7.300e-24 56-79 PR00196B 10.68 4.808e-16 96-113 PR00196A 11.16 6.236e-14 128-151 PR00196E 9.19 1.000e-12 127-148 PR00196G 11.72 5.829e-11 171-185 PR00196C 10.36 7.913e-11 51-73 PR00196C 10.36 8.750e-10 282-304
1217	PR00202	ANNEXIN TYPE VI SIGNATURE	PR00202D 5.58 6.793e-14 123-145 PR00202G 8.01 5.545e-13 127-154 PR00202E 13.00 8.740e-11 206-233 PR00202B 11.44 2.782e-10 127-151 PR00202B 11.44 5.206e-09 55-79
1217	PR00197	ANNEXIN TYPE I SIGNATURE	PR00197E 11.89 1.794e-32 206-233 PR00197B 7.56 4.960e-29 56-79 PR00197D 7.50 3.000e-26 123-145 PR00197A 8.68 7.577e-20 7-23 PR00197C 7.50 1.000e-19 96-113 PR00197F 9.03 7.128e-10 127-148 PR00197D 7.50 6.250e-09 51-73

SEQ ID NO:	Database entry ID	Description	Results*
1217	PR00200	ANNEXIN TYPE IV SIGNATURE	PR00200E 10.00 5.030e-19 123-145 PR00200G 9.43 5.546e-14 127-154 PR00200B 7.39 4.653e-11 128-151 PR00200F 13.72 7.094e-11 206-233 PR00200B 7.39 4.857e-10 56-79 PR00200H 13.68 9.663e-10 171-185 PR00200E 10.00 8.842e-09 282-304
1217	PR00199	ANNEXIN TYPE III SIGNATURE	PR00199D 5.65 9.297e-17 123-145 PR00199F 16.19 4.391e-15 206-233 PR00199B 6.86 2.915e-13 56-79 PR00199B 6.86 1.265e-11 128-151 PR00199G 9.09 4.351e-11 128-154 PR00199D 5.65 3.641e-09 51-73 PR00199C 13.84 9.571e-09 96-113
1218	BL00223	Annexins repeat proteins domain proteins.	BL00223A 15.59 1.000e-33 119-153 BL00223A 15.59 1.435e-16 47-81 BL00223C 24.79 3.928e-15 106-161
1218	PR00196	ANNEXIN FAMILY SIGNATURE	PR00196C 10.36 3.571e-25 123-145 PR00196A 11.16 7.300e-24 56-79 PR00196B 10.68 4.808e-16 96-113 PR00196A 11.16 6.235e-14 128-151 PR00196E 9.19 1.000e-12 127-148 PR00196G 11.72 5.829e-11 171-185 PR00196C 10.36 7.913e-11 51-73
1218	PR00201	ANNEXIN TYPE V SIGNATURE	PR00201D 10.49 8.729e-14 123-145 PR00201A 6.05 1.976e-13 56-79 PR00201G 11.02 3.847e-12 127-154 PR00201A 6.05 8.241e-12 128-151 PR00201H 12.04 4.889e-10 171-185
1218	PR00198	ANNEXIN TYPE II SIGNATURE	PR00198D 7.65 7.787e-21 123-145 PR00198B 8.71 3.880e-17 56-79 PR00198C 14.32 2.688e-11 96-113 PR00198G 8.09 7.033e-10 127-148
1218	PR00200	ANNEXIN TYPE IV SIGNATURE	PR00200E 10.00 5.030e-19 123-145 PR00200G 9.43 5.546e-14 127-154 PR00200B 7.39 4.653e-11 128-151 PR00200B 7.39 4.857e-10 56-79 PR00200H 13.68 9.663e-10 171-185
1218	PR00202	ANNEXIN TYPE VI SIGNATURE	PR00202D 5.58 6.793e-14 123-145 PR00202G 8.01 5.545e-13 127-154 PR00202B 11.44 2.782e-10 127-151 PR00202B 11.44 5.206e-09 55-79
1218	PR00197	ANNEXIN TYPE I SIGNATURE	PR00197B 7.56 4.960e-29 56-79 PR00197D 7.50 3.000e-26 123-145 PR00197A 8.68 7.577e-20 7-23 PR00197C 7.50 1.000e-19 96-113 PR00197F 9.03 7.128e-10 127-148 PR00197D 7.50 6.250e-09 51-73

SEQ ID NO:	Database entry ID	Description	Results*
1218	PR00199	ANNEXIN TYPE III SIGNATURE	PR00199D 5.65 9.297e-17 123-145 PR00199B 6.86 2.915e-13 56-79 PR00199B 6.86 1.265e-11 128-151 PR00199G 9.09 4.351e-11 128-154 PR00199D 5.65 3.641e-09 51-73 PR00199C 13.84 9.571e-09 96-113
1221	BL00086	Cytochrome P450 cysteine heme-iron ligand proteins.	BL00086 20.87 8.615e-27 423-455
1221	PR00465	E-CLASS P450 GROUP IV SIGNATURE	PR00465F 13.37 8.468e-12 393-412
1221	PR00359	B-CLASS P450 SIGNATURE	PR00359I 11.13 7.261e-11 433-445
1221	PR00463	E-CLASS P450 GROUP I SIGNATURE	PR00463I 15.02 9.571e-21 433-457 PR00463G 18.24 6.760e-19 388-413 PR00463E 17.37 6.595e-17 304-331 PR00463F 17.63 7.568e-12 347-366 PR00463B 17.50 7.692e-12 79-101 PR00463D 14.02 8.875e-11 284-302 PR00463C 12.85 6.932e-10 171-190
1221	PR00464	E-CLASS P450 GROUP II SIGNATURE	PR00464G 12.41 2.588e-12 398-414 PR00464E 18.28 3.077e-10 342-363 PR00464I 14.64 3.106e-10 433-457 PR00464H 13.32 4.635e-09 420-434 PR00464D 17.40 5.787e-09 313-331 PR00464C 18.84 5.808e-09 284-313
1221	PR00385	P450 SUPERFAMILY SIGNATURE	PR00385E 12.66 9.100e-14 433-445 PR00385A 14.97 5.696e-13 295-313 PR00385B 10.22 6.400e-09 313-327
1221	PR00408	MITOCHONDRIAL P450 SIGNATURE	PR00408D 15.44 6.831e-09 295-313
1222	PR00385	P450 SUPERFAMILY SIGNATURE	PR00385A 14.97 5.696e-13 295-313
1222	PR00463	E-CLASS P450 GROUP I SIGNATURE	PR00463B 17.50 7.692e-12 79-101 PR00463D 14.02 8.875e-11 284-302 PR00463C 12.85 6.932e-10 171-190
1222	PR00464	E-CLASS P450 GROUP II SIGNATURE	PR00464C 18.84 5.808e-09 284-313
1222	PR00408	MITOCHONDRIAL P450 SIGNATURE	PR00408D 15.44 6.831e-09 295-313
1223	BL00477	Alpha-2-macroglobulin family thiolester region proteins.	BL00477A 13.50 9.182e-19 70-99
1225	BL00500	Thymosin beta-4 family proteins.	BL00500 9.77 2.565e-28 2-42
1227	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320A 16.74 4.971e-10 231-246 PR00320C 13.01 8.200e-10 231-246 PR00320B 12.19 9.486e-10 231-246 PR00320B 12.19 3.475e-09 188-203 PR00320B 12.19 4.600e-09 315-330 PR00320C 13.01 4.900e-09 315-330
1227	PR00319	BETA G-PROTEIN (TRANSDUCIN)	PR00319B 11.47 9.143e-09 315-

SEQ ID NO:	Database entry ID	Description	Results*
		SIGNATURE	330
1227	BL00678	Trp-Asp (WD) repeat proteins proteins.	BL00678 9.67 1.600e-10 233-244 BL00678 9.67 1.000e-08 317-328
1236	PF00580	UvrD/REP helicase.	PF00580D 13.15 8.920e-13 670-684 PF00580E 13.89 2.800e-11 867-886 PF00580F 8.62 9.438e-10 913-926
1237	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 1.000e-09 64-78 PR00019A 11.19 8.000e-09 90-104
1238	PD00126	PROTEIN REPEAT DOMAIN TPR NUCLEA.	PD00126A 22.53 5.500e-10 229-250
1243	DM01206	CORONAVIRUS NUCLEOCAPSID PROTEIN.	DM01206B 10.69 4.759e-09 464-484
1243	BL00315	Dehydrins proteins.	BL00315A 9.35 1.000e-08 389-417
1245	BL00972	Ubiquitin carboxyl-terminal hydrolases family 2 proteins.	BL00972A 11.93 8.054e-15 191-209
1246	PF00023	Ank repeat proteins.	PF00023A 16.03 9.500e-12 347-363 PF00023A 16.03 8.500e-10 283-299 PF00023A 16.03 8.875e-10 184-200
1246	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 1.989e-13 217-272 PF00791B 28.49 6.044e-13 117-172 PF00791B 28.49 4.316e-12 184-239 PF00791B 28.49 9.432e-12 250-305 PF00791B 28.49 6.243e-10 84-139 PF00791C 20.98 4.971e-09 98-137
1246	BL00906	Uroporphyrinogen decarboxylase proteins.	BL00906D 24.33 7.750e-09 212-256
1248	BL00415	Synapsins proteins.	BL00415Q 2.23 8.297e-09 13-49
1250	BL01113	C1q domain proteins.	BL01113B 18.26 2.500e-13 841-877
1252	BL01248	Laminin-type EGF-like (LE) domain proteins.	BL01248 11.02 7.171e-12 258-271 BL01248 11.02 7.943e-12 325-338
1252	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011D 14.03 7.000e-17 376-395 PR00011A 14.06 1.000e-14 376-395 PR00011B 13.08 5.167e-14 376-395 PR00011C 24.25 8.468e-14 395-424 PR00011D 14.03 9.739e-09 249-268
1253	BL00164	Enolase proteins.	BL00164A 11.58 2.800e-28 41-64
1253	PR00148	ENOLASE SIGNATURE	PR00148A 10.11 1.783e-18 44-59
1255	BL01153	NOL1/NOP2/sun family proteins.	BL01153D 19.69 8.322e-14 102-128 BL01153C 13.67 6.507e-10 51-65
1256	BL00892	HIT family proteins.	BL00892B 16.86 1.000e-20 130-154 BL00892A 18.17 6.657e-20 64-95
1256	PR00332	HISTIDINE TRIAD FAMILY SIGNATURE	PR00332B 13.62 3.000e-16 76-95 PR00332C 7.37 4.600e-14 143-154 PR00332A 10.15 7.375e-12 55-72
1257	PF00791	Domain present in ZO-1 and Unc5-like netrin	PF00791B 28.49 4.146e-10 73-

SEQ ID NO:	Database entry ID	Description	Results ^a
		receptors.	128
1258	BL00615	C-type lectin domain proteins.	BL00615B 12.25 5.200e-12 166-180
1259	BL00071	Glyceraldehyde 3-phosphate dehydrogenase proteins.	BL00071B 21.70 1.000e-40 80-126 BL00071C 11.81 1.000e-40 146-181 BL00071D 19.39 3.118e-25 184-239 BL00071E 11.48 4.600e-24 308-329 BL00071A 5.81 2.607e-14 5-17
1259	PR00078	GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE SIGNATURE	PR00078B 7.45 3.250e-24 146-165 PR00078D 11.49 2.800e-21 231-249 PR00078E 10.50 6.211e-16 271-287 PR00078A 10.38 1.000e-15 111-125 PR00078C 15.99 6.211e-11 173-190
1262	PR00926	MITOCHONDRIAL CARRIER PROTEIN SIGNATURE	PR00926F 17.75 2.688e-10 15-38 PR00926D 10.53 6.625e-10 21-40
1262	PR00927	ADENINE NUCLEOTIDE TRANSLOCATOR 1 SIGNATURE	PR00927E 14.93 6.143e-10 44-66 PR00927B 14.66 9.870e-10 265-287 PR00927A 14.66 5.685e-09 46-68
1262	BL00215	Mitochondrial energy transfer proteins.	BL00215A 15.82 6.250e-17 13-38 BL00215A 15.82 1.600e-15 230-255 BL00215A 15.82 5.974e-13 108-133 BL00215B 10.44 7.600e-09 275-288
1263	PR00654	ANGIOTENSINOGEN SIGNATURE	PR00654A 15.64 1.540e-26 23-44 PR00654D 10.48 3.538e-26 153-175 PR00654F 15.16 8.071e-26 255-275 PR00654E 9.81 2.241e-25 194-215 PR00654C 9.50 5.500e-21 115-135
1263	BL00284	Serpins proteins.	BL00284C 28.56 9.514e-21 254-296 BL00284E 19.15 9.710e-16 439-464 BL00284A 15.64 8.147e-14 113-137 BL00284D 16.34 1.837e-12 361-388 BL00284B 17.99 7.500e-11 229-250
1264	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 8.071e-17 34-57
1265	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 7.600e-16 34-57
1266	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 2.800e-16 31-54
1267	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 9.400e-16 34-57
1268	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290B 13.17 4.000e-21 282-300 BL00290A 20.89 4.600e-16 34-57 BL00290A 20.89 2.421e-13 225-248
1269	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 4.600e-16 34-57
1271	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 8.071e-17 34-57
1272	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 4.600e-16 34-57
1273	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 4.600e-16 34-57

SEQ ID NO:	Database entry ID	Description	Results*
		histocompatibility complex proteins.	
1274	PF00651	BTB (also known as BR-C/Tik) domain proteins.	PF00651 15.00 2.500e-14 46-59
1274	DM00099	4 kw A55R REDUCTASE TERMINAL DIHYDROPTERIDINE.	DM00099B 14.73 6.063e-09 300-310
1274	PR00501	KELCH REPEAT SIGNATURE	PR00501A 8.25 7.750e-11 472-486 PR00501A 8.25 7.955e-09 328-342
1281	BL00972	Ubiquitin carboxyl-terminal hydrolases family 2 proteins.	BL00972A 11.93 3.919e-15 101-119 BL00972B 9.45 7.577e-10 180-190
1285	PF00789	Domain present in ubiquitin-regulatory proteins.	PF00789B 19.70 5.941e-09 213-234
1286	PF00789	Domain present in ubiquitin-regulatory proteins.	PF00789B 19.70 5.941e-09 259-280
1287	PR00625	DNAJ PROTEIN FAMILY SIGNATURE	PR00625A 12.84 3.000e-19 19-39 PR00625B 13.48 2.756e-17 47-68
1287	BL00636	Ni-dnaJ domain proteins.	BL00636A 8.07 7.600e-19 23-40 BL00636B 15.11 6.870e-15 47-68
1288	BL00518	Zinc finger, C3HC4 type (RING finger), proteins.	BL00518 12.23 5.714e-10 24-33
1289	PR00500	POLYCYSTIC KIDNEY DISEASE PROTEIN SIGNATURE	PR005001 9.22 1.107e-31 2810-2833 PR00500G 3.68 1.087e-30 2525-2548 PR00500H 17.80 1.107e-29 2662-2684 PR00500E 6.99 1.106e-27 2350-2370 PR00500F 9.44 1.108e-26 2483-2503
1289	PF00801	PKD domain proteins.	PF00801B 23.63 9.217e-26 1055-1083 PF00801A 13.49 6.276e-11 222-235 PF00801B 23.63 3.087e-10 719-747 PF00801B 23.63 6.609e-10 1652-1680
1291	BL00415	Synapsins proteins.	BL00415N 4.29 5.401e-09 136-180
1292	PD00930	PROTEIN GTPASE DOMAIN ACTIVATION.	PD00930B 33.72 2.800e-23 229-270 PD00930A 25.62 5.021e-12 125-151
1292	PF00620	GTPase-activator protein for Rho-like GTPases.	PF00620B 14.20 7.000e-12 178-195
1293	BL00023	Type II fibronectin collagen-binding domain proteins.	BL00023 24.31 8.043e-34 281-318 BL00023 24.31 5.320e-32 223-260 BL00023 24.31 5.800e-29 340-377
1293	BL00142	Neutral zinc metalloproteases, zinc-binding region proteins.	BL00142 8.38 1.000e-12 398-409
1293	PR00138	MATRIXIN SIGNATURE	PR00138D 16.56 5.500e-30 398-424 PR00138C 16.41 1.000e-29 164-193 PR00138B 15.82 5.875e-16 141-157 PR00138A 15.14 1.000e-15 94-108 PR00138E 6.01 8.472e-11 431-445
1293	BL00024	Hemopexin domain proteins.	BL00024C 22.98 1.000e-40 163-212 BL00024D 17.28 4.316e-36 392-424 BL00024B 21.53 7.545e-33 115-149 BL00024F 11.30 2.895e-18 486-507 BL00024A 11.49 3.667e-12 94-105

SEQ ID NO:	Database entry ID	Description	Results ^a
			BL00024G 13.31 4.857e-12 525-538 BL00024E 7.58 2.263e-10 431-445
1293	PR00480	ASTACIN FAMILY SIGNATURE	PR00480B 15.41 1.771e-09 393-412
1293	PR00013	FIBRONECTIN TYPE II REPEAT SIGNATURE	PR00013C 12.29 7.136e-19 372-388 PR00013C 12.29 6.464e-18 313-329 PR00013C 12.29 5.500e-16 255-271 PR00013B 14.75 4.375e-12 355-368 PR00013A 12.26 1.973e-11 344-354 PR00013A 12.26 6.595e-11 227-237 PR00013B 14.75 8.568e-10 296-309 PR00013A 12.26 2.167e-09 285-295 PR00013B 14.75 2.588e-09 238-251
1293	BL00546	Matrixins cysteine switch.	BL00546B 20.11 3.368e-40 164-208 BL00546C 16.41 6.400e-31 392-424 BL00546A 19.62 6.523e-19 74-104 BL00546E 10.23 7.947e-13 486-507 BL00546F 12.40 5.339e-09 525-538
1294	BL00023	Type II fibronectin collagen-binding domain proteins.	BL00023 24.31 8.043e-34 281-318 BL00023 24.31 5.320e-32 223-260 BL00023 24.31 5.800e-29 340-377
1294	BL00546	Matrixins cysteine switch.	BL00546B 20.11 3.368e-40 164-208 BL00546C 16.41 6.400e-31 392-424 BL00546A 19.62 6.523e-19 74-104
1294	PR00138	MATRIXIN SIGNATURE	PR00138D 16.56 5.500e-30 398-424 PR00138C 16.41 1.000e-29 164-193 PR00138B 15.82 5.875e-16 141-157 PR00138A 15.14 1.000e-15 94-108
1294	BL00142	Neutral zinc metalloproteinases, zinc-binding region proteins.	BL00142 8.38 1.000e-12 398-409
1294	BL00024	Hemopexin domain proteins.	BL00024C 22.98 1.000e-40 163-212 BL00024D 17.28 4.316e-36 392-424 BL00024B 21.53 7.545e-33 115-149 BL00024A 11.49 3.667e-12 94-105
1294	PR00480	ASTACIN FAMILY SIGNATURE	PR00480B 15.41 1.771e-09 393-412
1294	PR00013	FIBRONECTIN TYPE II REPEAT SIGNATURE	PR00013C 12.29 7.136e-19 372-388 PR00013C 12.29 6.464e-18 313-329 PR00013C 12.29 5.500e-16 255-271 PR00013B 14.75 4.375e-12 355-368 PR00013A 12.26 1.973e-11 344-354 PR00013A 12.26 6.595e-11 227-237 PR00013B 14.75 8.568e-10 296-309 PR00013A 12.26 2.167e-09 285-295 PR00013B 14.75 2.588e-09 238-251
1298	DM01354	kw TRANSCRIPTASE REVERSE II ORF2.	DM01354R 8.50 2.969e-22 2115-2145 DM01354S 11.61 1.692e-14

SEQ ID NO:	Database entry ID	Description	Results*
			2145-2166
1298	DM01688	2 POLY-IG RECEPTOR.	DM01688D 13.44 8.244e-09 1714-1737
1298	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 5.737e-10 1807-1817 DM00179 13.97 7.158e-10 1077-1087 DM00179 13.97 9.053e-10 759-769 DM00179 13.97 9.053e-10 1328-1338 DM00179 13.97 4.130e-09 574-584 DM00179 13.97 4.130e-09 1431-1441 DM00179 13.97 6.870e-09 1713-1723 DM00179 13.97 7.652e-09 850-860 DM00179 13.97 8.435e-09 2089-2099
1298	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 8.909e-13 623-647 BL00240B 24.70 1.209e-10 1126-1150 BL00240B 24.70 4.558e-10 124-148 BL00240B 24.70 6.442e-10 529-553 BL00240B 24.70 4.255e-09 1222-1246 BL00240B 24.70 8.468e-09 995-1019
1298	PD02327	GLYCOPROTEIN ANTIGEN PRECURSOR IMMUNOGLO.	PD02327B 19.84 8.091e-09 1126-1148 PD02327B 19.84 9.318e-09 1222-1244
1298	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 1.200e-10 1610-1643 PD02870B 18.83 7.400e-10 2081-2114 PD02870B 18.83 7.800e-10 1069-1102 PD02870B 18.83 5.213e-09 1423-1456 PD02870B 18.83 6.649e-09 67-100 PD02870B 18.83 7.989e-09 1518-1551 PD02870D 15.74 8.564e-09 566-601 PD02870B 18.83 9.521e-09 286-319 PD02870B 18.83 9.904e-09 1258-1291
1299	BL00888	Cyclic nucleotide-binding domain proteins.	BL00888B 14.79 4.706e-18 372-396 BL00888A 18.03 1.000e-08 354-371
1301	PF00615	Regulator of G protein signalling domain proteins.	PF00615B 16.25 9.625e-16 73-90 PF00615C 10.06 9.206e-12 150-164
1302	BL00766	Tetrahydrofolate dehydrogenase/cyclohydrolase proteins.	BL00766E 13.78 9.625e-39 191-228 BL00766C 25.86 4.375e-31 77-125 BL00766D 17.05 5.966e-25 152-182
1302	PR00085	TETRAHYDROFOLATE DEHYDROGENASE/CYCLOHYDROLASE FAMILY SIGNATURE	PR00085E 15.79 7.000e-26 151-181 PR00085G 10.74 1.865e-22 208-227 PR00085C 15.23 6.182e-21 47-69 PR00085D 15.02 2.688e-20 92-113 PR00085F 9.77 6.595e-15 191-208
1303	BL00180	Glutamine synthetase proteins.	BL00180E 17.60 1.000e-40 154-206 BL00180D 13.26 2.174e-24 119-141 BL00180F 10.05 6.211e-17 218-231 BL00180G 10.20

SEQ ID NO:	Database entry ID	Description	Results*
			8.435e-17 307-322 BL00180C 12.14 4.600e-14 102-112 BL00180B 18.03 4.971e-14 68-87 BL00180A 13.20 5.065e-14 32-45
1304	BL00180	Glutamine synthetase proteins.	BL00180F 10.03 6.750e-15 49-62
1306	BL01131	Ribosomal RNA adenine dimethylases proteins.	BL01131A 26.62 1.000e-08 77-123
1308	DM00191	w SPAC8A4.04C RESISTANCE SPAC8A4.05C DAUNORUBICIN.	DM00191A 8.16 5.440e-09 36-49
1309	DM00191	w SPAC8A4.04C RESISTANCE SPAC8A4.05C DAUNORUBICIN.	DM00191A 8.16 5.440e-09 61-74
1310	PF00606	Herpesviral Glycoprotein B.	PF00606I 20.74 7.894e-09 264-316
1310	BL01219	Ammonium transporters proteins.	BL01219D 11.63 2.957e-10 217-241 BL01219F 15.24 8.809e-09 335-360
1310	PR00342	RHESUS BLOOD GROUP PROTEIN SIGNATURE	PR00342G 8.18 1.458e-19 220-239 PR00342B 11.09 8.657e-13 61-79 PR00342D 8.46 2.857e-12 128-145 PR00342I 4.99 6.016e-12 285-299 PR00342H 7.61 6.927e-11 250-273 PR00342C 10.10 4.770e-10 90-108 PR00342E 14.49 5.950e-10 151-175 PR00342F 7.02 1.556e-09 185-201 PR00342J 8.97 7.940e-09 308-327 PR00342L 7.61 9.600e-09 398-424
1311	PR00209	ALPHA/BETA GLIADIN FAMILY SIGNATURE	PR00209B 4.88 9.080e-11 80-99 PR00209B 4.88 9.967e-10 86-105
1311	DM00406	GLIADIN.	DM00406 7.73 1.400e-09 86-99
1311	PR00501	KELCH REPEAT SIGNATURE	PR00501B 18.88 8.342e-09 440-455
1312	PR00528	GLUCOCORTICOID RECEPTOR SIGNATURE	PR00528F 9.13 9.063e-09 31-51
1313	PF00622	Domain in SP1a and the RYanodine Receptor.	PF00622C 12.62 6.625e-13 759-773
1313	BL00518	Zinc finger, C3HC4 type (RING finger), proteins.	BL00518 12.23 6.571e-10 30-39 BL00518 12.23 1.667e-09 356-365
1314	BL00420	Speract receptor repeat proteins domain proteins.	BL00420B 22.67 2.824e-25 37-92 BL00420C 11.90 9.250e-12 122-133
1314	PR00258	SPERACT RECEPTOR SIGNATURE	PR00258D 14.41 6.333e-11 98-113 PR00258B 9.63 7.474e-11 52-64 PR00258E 13.33 1.750e-09 121-134 PR00258C 9.05 5.167e-09 67-78
1315	PR00080	ALCOHOL DEHYDROGENASE SUPERFAMILY SIGNATURE	PR00080A 9.32 8.548e-10 122-134
1315	BL00766	Tetrahydrofolate dehydrogenase/cyclohydrolase proteins.	BL00766C 25.86 7.632e-09 20-68
1315	PR00081	GLUCOSE/RIBITOL DEHYDROGENASE FAMILY SIGNATURE	PR00081A 10.53 2.452e-13 41-59 PR00081C 15.13 9.229e-09 167-184
1317	BL00263	Natriuretic peptides proteins.	BL00263 11.87 5.909e-22 129-147

SEQ ID NO:	Database entry ID	Description	Results*
1317	PR00711	ATRIAL NATRIURETIC PEPTIDE SIGNATURE	PR00711G 11.75 1.113e-30 128-151 PR00711B 10.71 7.545e-24 32-51 PR00711D 7.91 1.000e-22 72-90 PR00711F 3.17 2.463e-22 109-128 PR00711C 5.66 7.818e-22 51-70 PR00711E 6.39 1.000e-21 92-109 PR00711A 12.00 9.769e-20 11-30
1317	PR00713	C-TYPE NATRIURETIC PEPTIDE SIGNATURE	PR00713C 14.14 1.370e-13 130-146
1317	PR00710	NATRIURETIC PEPTIDE FAMILY SIGNATURE	PR00710A 10.90 3.250e-14 127-137 PR00710B 11.08 1.391e-12 136-146
1317	PR00712	BRAIN NATRIURETIC PEPTIDE SIGNATURE	PR00712D 10.52 4.109e-12 128-139 PR00712E 10.62 7.231e-10 138-152
1318	BL00609	Glycosyl hydrolases family 32 proteins.	BL00609C 13.27 9.270e-11 249-261
1318	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 6.538e-16 757-773 BL01187B 12.04 7.750e-14 610-626 BL01187B 12.04 8.200e-14 651-667 BL01187B 12.04 2.029e-10 523-539 BL01187A 9.98 7.429e-10 591-603
1318	PR00907	THROMBOMODULIN SIGNATURE	PR00907B 11.29 6.301e-11 753-770 PR00907B 11.29 2.636e-10 647-664 PR00907B 11.29 3.524e-09 519-536 PR00907G 11.63 4.243e-09 651-678
1318	BL01177	Anaphylatoxin domain proteins.	BL01177C 17.39 8.286e-09 517-536
1318	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010C 11.16 1.429e-09 762-773 PR00010C 11.16 8.500e-09 528-539
1318	BL00022	EGF-like domain proteins.	BL00022B 7.54 1.000e-08 619-626
1319	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237E 13.03 4.000e-10 26-50
1320	BL00125	Serine/threonine specific protein phosphatases proteins.	BL00125D 33.11 9.719e-35 23-78
1320	PR00114	SERINE/THREONINE PHOSPHATASE FAMILY SIGNATURE	PR00114F 17.51 4.706e-16 39-60 PR00114G 17.20 5.421e-12 61-78
1321	BL00453	EKBP-type peptidyl-prolyl cis-trans isomerase proteins.	BL00453B 23.86 6.538e-26 281-315 BL00453A 15.57 8.364e-12 249-264 BL00453C 9.72 3.250e-11 323-336
1321	PR00280	CHANNEL FORMING COLICIN SIGNATURE	PR00280A 11.09 8.227e-09 284-300
1322	PR00497	NEUTROPHIL CYTOSOL FACTOR P40 SIGNATURE	PR00497A 6.92 8.261e-09 310-328
1322	BL50002	Src homology 3 (SH3) domain proteins profile.	BL50002B 15.18 9.500e-09 45-59
1323	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 8.269e-16 34-56
1323	BL01115	GTP-binding nuclear protein ran proteins.	BL01115A 10.22 2.474e-09 34-78
1323	PR00300	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT SIGNATURE	PR00300A 9.56 6.260e-09 36-55

SEQ ID NO:	Database entry ID	Description	Results*
1323	BL00567	Phosphoribulokinase proteins.	BL00567A 10.66 9.100e-09 35-54
1324	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 9.847e-10 314-333
1324	PF00242	DNA polymerase (viral) N-terminal domain proteins.	PF00242G 13.52 6.276e-09 748-788
1324	BL00107	Protein kinases ATP-binding region proteins.	BL00107B 13.31 1.000e-11 381-397 BL00107A 18.39 8.091e-09 314-345
1325	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 9.847e-10 314-333
1325	PF00242	DNA polymerase (viral) N-terminal domain proteins.	PF00242G 13.52 6.276e-09 721-761
1325	BL00107	Protein kinases ATP-binding region proteins.	BL00107B 13.31 1.000e-11 381-397 BL00107A 18.39 8.091e-09 314-345
1326	BL00472	Small cytokines (intercrine/chemokine) C-C subfamily signatur.	BL00472C 20.76 8.225e-09 50-87
1327	PR00705	PAPAIN CYSTEINE PROTEASE (C1) FAMILY SIGNATURE	PR00705A 10.55 8.667e-13 114-130 PR00705B 10.22 2.385e-10 293-304
1327	BL00139	Eukaryotic thiol (cysteine) proteases cysteine proteins.	BL00139D 9.24 8.125e-17 312-329 BL00139C 9.23 2.800e-10 292-302 BL00139B 10.19 7.600e-10 157-166 BL00139A 10.29 2.723e-09 114-124
1328	PR00705	PAPAIN CYSTEINE PROTEASE (C1) FAMILY SIGNATURE	PR00705A 10.55 8.667e-13 155-171 PR00705B 10.22 2.385e-10 334-345
1328	BL00139	Eukaryotic thiol (cysteine) proteases cysteine proteins.	BL00139D 9.24 8.125e-17 353-370 BL00139C 9.23 2.800e-10 333-343 BL00139B 10.19 7.600e-10 198-207 BL00139A 10.29 2.723e-09 155-165
1330	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270A 17.22 7.443e-10 129-169 PD01270A 17.22 7.387e-09 36-76
1332	BL00412	Neuromodulin (GAP-43) proteins.	BL00412D 16.54 6.772e-10 250-301
1332	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 2.068e-09 751-784
1333	BL00232	Cadherins extracellular repeat proteins domain proteins.	BL00232B 32.79 8.594e-35 151-199 BL00232B 32.79 5.579e-22 260-308 BL00232A 27.72 1.000e-20 57-90 BL00232C 10.65 3.613e-14 258-276 BL00232B 32.79 4.872e-11 377-425 BL00232C 10.65 3.211e-09 480-498
1333	DM01724	kw ALLERGEN POLLEN CIM1 HOL-LI.	DM01724 8.14 9.113e-10 698-718 DM01724 8.14 6.803e-09 694-714
1333	PR00205	CADHERIN SIGNATURE	PR00205B 11.39 4.545e-15 258-276 PR00205A 14.73 5.600e-09 183-199 PR00205B 11.39 8.017e-09 480-498
1335	BI.00214	Cytosolic fatty-acid binding proteins.	BL00214B 26.51 9.000e-29 47-92 BL00214A 21.17 1.000e-24 6-32
1335	PR00178	FATTY ACID-BINDING PROTEIN	PR00178C 20.54 3.864e-25 65-93

SEQ ID NO:	Database entry ID	Description	Results*
		SIGNATURE	PR00178A 15.07 7.188e-23 7-28 PR00178D 13.52 6.170e-12 111-130
1336	PR00452	SH3 DOMAIN SIGNATURE	PR00452B 11.65 8.250e-09 509-525
1338	BL00972	Ubiquitin carboxyl-terminal hydrolases family 2 proteins.	BL00972A 11.93 8.759e-17 112-130 BL00972D 22.55 8.116e-12 354-379 BL00972B 9.45 7.088e-09 193-203
1340	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243I 31.77 3.155e-09 1-44
1340	PR00633	CHROMOSOME CONDENSATION REGULATOR RCC1 SIGNATURE	PR00633E 12.18 4.682e-10 182-199 PR00633G 13.71 1.667e-09 185-204 PR00633H 15.10 3.963e-09 244-266
1340	BL00625	Regulator of chromosome condensation (RCC1) proteins.	BL00625B 17.69 5.219e-15 179-213 BL00625B 17.69 9.194e-14 343-377 BL00625A 16.21 4.405e-12 185-214 BL00625A 16.21 5.500e-12 129-158 BL00625A 16.21 7.203e-12 349-378 BL00625B 17.69 5.778e-10 123-157 BL00625B 17.69 5.034e-09 285-319
1342	BL00476	Fatty acid desaturases family 1 proteins.	BL00476F 12.75 6.551e-09 45-90
1345	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320A 16.74 9.690e-11 292-307 PR00320B 12.19 4.343e-10 292-307 PR00320C 13.01 7.840e-10 292-307
1345	BL00678	Tip-Asp (WD) repeat proteins proteins.	BL00678 9.67 1.000e-09 294-305
1345	BL00225	Crystallins beta and gamma 'Greek key' motif proteins.	BL00225B 18.06 7.273e-22 6-41 BL00225B 18.06 5.673e-14 97-132 BL00225A 13.82 7.218e-09 61-82
1350	PD01823	PROTEIN INTERGENIC REGION ABC1 PRECURSOR MITOCHONDRION T.	PD01823D 16.66 3.093e-15 21-42 PD01823E 9.30 5.909e-15 75-88
1352	BL00540	Ferritin iron-binding regions proteins.	BL00540A 15.06 1.000e-40 9-50 BL00540B 18.82 1.000e-40 100-155 BL00540C 13.00 7.500e-15 165-177
1353	PR00294	STREPTOMYCES SUBTILISIN INHIBITOR SIGNATURE	PR00294A 10.44 6.444e-10 159-186
1353	DM00191	w SPAC8A4.04C RESISTANCE SPAC8A4.05C DAUNORUBICIN.	DM00191D 13.94 7.167e-10 206-245
1356	BL00428	Cell cycle proteins fisW / rodA / spoVE proteins.	BL00428A 14.30 3.613e-09 91-110
1359	BL00142	Neutral zinc metalloproteinases, zinc-binding region proteins.	BL00142 8.38 7.188e-10 389-400
1359	PD01719	PRECURSOR GLYCOPROTEIN SIGNAL RE.	PD01719A 12.89 7.983e-16 550-578 PD01719B 9.30 1.750e-09 877-885 PD01719A 12.89 3.000e-09 1006-1034
1359	PR00480	ASTACIN FAMILY SIGNATURE	PR00480B 15.41 3.186e-09 384-403
1360	DM00191	w SPAC8A4.04C RESISTANCE SPAC8A4.05C DAUNORUBICIN.	DM00191D 13.94 6.330e-11 232-271 DM00191D 13.94 7.728e-11 48-87 DM00191D 13.94 5.000e-10 112-151 DM00191D 13.94

SEQ ID NO:	Database entry ID	Description	Results*
			5.667e-10 59-98 DM00191D 13.94 5.667e-10 123-162 DM00191D 13.94 6.583e-10 56-95 DM00191D 13.94 8.417e-10 280-319 DM00191D 13.94 8.917e-10 192-231 DM00191D 13.94 1.391e-09 224-263 DM00191D 13.94 2.409e-09 208-247 DM00191D 13.94 4.835e-09 120-159 DM00191D 13.94 5.304e-09 149-188 DM00191D 13.94 5.461e-09 211-250 DM00191D 13.94 6.322e-09 80-119 DM00191D 13.94 7.652e-09 243-282 DM00191D 13.94 8.513e-09 216-255 DM00191D 13.94 9.452e-09 177-216
1360	PF00624	Flocculin repeat proteins.	PF00624J 6.21 3.496e-11 237-292 PF00624J 6.21 6.597e-11 53-108 PF00624J 6.21 4.121e-10 253-308 PF00624J 6.21 5.718e-10 141-196 PF00624F 11.04 1.508e-09 50-86 PF00624J 6.21 3.163e-09 101-156 PF00624J 6.21 3.233e-09 165-220 PF00624J 9.10 5.181e-09 140-170 PF00624F 11.04 6.008e-09 130-166 PF00624J 6.21 6.093e-09 125-180 PF00624J 6.21 6.163e-09 221-276 PF00624G 10.91 6.806e-09 45-100 PF00624G 10.91 7.169e-09 181-236 PF00624G 10.91 7.387e-09 221-276 PF00624J 6.21 8.674e-09 197-252 PF00624J 6.21 8.884e-09 117-172 PF00624J 6.21 8.884e-09 213-268 PF00624J 6.21 9.512e-09 55-110
1360	BL00115	Eukaryotic RNA polymerase II heptapeptide repeat proteins.	BL00115Z 3.12 6.163e-10 22-71 BL00115Z 3.12 7.618e-09 36-85 BL00115Z 3.12 9.603e-09 241-290
1363	PF00023	Ank repeat proteins.	PF00023A 16.03 1.321e-09 110-126
1363	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 7.527e-13 110-165 PF00791B 28.49 6.119e-09 77-132 PF00791C 20.98 7.529e-09 91-130
1366	PF00168	C2 domain proteins.	PF00168C 27.49 9.250e-17 320-346
1366	PR00399	SYNAPTOTAGMIN SIGNATURE	PR00399A 9.52 1.844e-14 148-164 PR00399C 12.82 8.071e-14 218-234 PR00399B 14.27 2.853e-13 163-177 PR00399D 14.48 1.871e-11 238-249
1366	PR00360	C2 DOMAIN SIGNATURE	PR00360B 13.61 8.759e-12 337-351 PR00360A 14.59 3.903e-10 308-321 PR00360B 13.61 4.789e-10 203-217

SEQ ID NO:	Database entry ID	Description	Results*
1367	PF00168	C2 domain proteins.	PF00168C 27.49 9.250e-17 320-346
1367	PR00399	SYNAPTOTAGMIN SIGNATURE	PR00399A 9.52 1.844e-14 148-164 PR00399C 12.82 8.071e-14 218-234 PR00399B 14.27 2.853e-13 163-177 PR00399D 14.48 1.871e-11 238-249
1367	PR00360	C2 DOMAIN SIGNATURE	PR00360B 13.61 8.759e-12 337-351 PR00360A 14.59 3.903e-10 308-321 PR00360B 13.61 4.789e-10 203-217
1368	BL00086	Cytochrome P450 cysteine heme-iron ligand proteins.	BL00086 20.87 1.857e-20 444-476
1368	PR00464	E-CLASS P450 GROUP II SIGNATURE	PR00464I 14.64 4.375e-17 454-478 PR00464A 20.47 9.591e-16 130-151 PR00464C 18.84 1.000e-15 305-334 PR00464D 17.40 6.250e-15 334-352 PR00464H 13.32 8.941e-15 441-455 PR00464F 15.23 9.654e-13 403-419 PR00464B 20.41 1.844e-12 186-205 PR00464E 18.28 7.907e-12 363-384 PR00464G 12.41 8.412e-12 419-435
1368	PR00465	E-CLASS P450 GROUP IV SIGNATURE	PR00465H 17.76 6.586e-10 454-473
1368	PR00385	P450 SUPERFAMILY SIGNATURE	PR00385A 14.97 1.346e-12 316-334 PR00385B 10.22 4.130e-11 334-348 PR00385D 13.11 7.857e-10 445-455 PR00385E 12.66 9.438e-10 454-466
1368	PR00463	E-CLASS P450 GROUP I SIGNATURE	PR00463G 18.24 3.605e-14 409-434 PR00463E 17.37 4.814e-11 325-352 PR00463I 15.02 5.574e-09 454-478 PR00463H 12.41 7.158e-09 444-455
1370	BL00218	Amino acid permeases proteins.	BL00218D 21.49 9.757e-11 263-308
1371	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 3.288e-09 35-50
1372	PR00380	KINESIN HEAVY CHAIN SIGNATURE	PR00380A 14.18 4.086e-22 84-106 PR00380C 13.18 5.286e-17 240-259 PR00380D 9.93 7.698e-17 290-312 PR00380B 12.64 7.805e-14 207-225
1372	BL00411	Kinesin motor domain proteins.	BL00411G 21.39 7.750e-25 241-283 BL00411C 15.04 2.500e-22 84-106 BL00411H 15.66 8.235e-16 289-320 BL00411E 10.43 9.129e-16 135-154 BL00411F 14.77 9.795e-16 198-223 BL00411D 12.13 5.909e-09 114-125
1373	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 1.915e-09 590-605
1373	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 5.800e-12 158-172 PR00019B 11.36 1.000e-10 130-144 PR00019A 11.19 8.826e-10 133-147 PR00019B 11.36

SEQ ID NO:	Database entry ID	Description	Results*
			7.120e-09 106-120
1373	PR00500	POLYCYSTIC KIDNEY DISEASE PROTEIN SIGNATURE	PR00500B 7.74 7.821e-09 250-271
1374	BL00411	Kinesin motor domain proteins.	BL00411H 15.66 7.811e-22 79-110 BL00411G 21.39 8.683e-22 31-73
1374	PR00380	KINESIN HEAVY CHAIN SIGNATURE	PR00380C 13.18 2.385e-16 30-49 PR00380D 9.93 3.739e-16 80-102
1376	BL00790	Receptor tyrosine kinase class V proteins.	BL00790E 29.58 6.667e-12 767-815
1376	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010C 11.16 5.636e-10 423-434 PR00010C 11.16 8.071e-09 148-159
1376	BL00022	EGF-like domain proteins.	BL00022B 7.54 8.200e-09 427-434
1376	PR00907	THROMBOMODULIN SIGNATURE	PR00907B 11.29 7.312e-10 224-241 PR00907G 11.63 5.297e-09 62-89 PR00907B 11.29 8.354e-09 98-115 PR00907B 11.29 9.451e-09 334-351
1376	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 5.235e-15 62-78 BL01187B 12.04 5.765e-15 418-434 BL01187B 12.04 3.000e-12 143-159 BL01187B 12.04 7.333e-12 297-313 BL01187B 12.04 7.000e-11 338-354 BL01187B 12.04 4.857e-10 378-394 BL01187B 12.04 5.886e-10 102-118 BL01187A 9.98 6.571e-10 321-333 BL01187A 9.98 5.125e-09 126-138 BL01187A 9.98 9.625e-09 362-374
1377	BL00048	Protamine P1 proteins.	BL00048 6.39 4.038e-09 396-423
1381	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 8.500e-27 342-373
1381	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 5.412e-12 342-361
1381	PD00306	PROTEIN GLYCOPROTEIN PRECURSOR RE.	PD00306A 10.26 6.143e-09 25-39
1382	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 6.036e-09 48-61
1388	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 7.261e-10 69-83 PR00019B 11.36 4.600e-09 66-80
1392	BL00018	EF-hand calcium-binding domain proteins.	BL00018 7.41 6.870e-09 42-55
1396	BL00790	Receptor tyrosine kinase class V proteins.	BL00790B 21.59 1.000e-40 61-113 BL00790C 16.65 1.000e-40 165-219 BL00790K 9.30 1.000e-40 657-711 BL00790Q 15.61 1.000e-40 855-904 BL00790O 7.68 5.929e-39 797-830 BL00790G 22.05 5.114e-36 376-420 BL00790R 16.20 7.469e-36 951-995 BL00790E 29.58 7.250e-35 273-321 BL00790J 14.21 8.200e-33 605-645 BL00790N 13.25 1.214e-31 763-790 BL00790I 20.01 1.931e-29 501-532 BL00790D 12.41 2.500e-27

SEQ ID NO:	Database entry ID	Description	Results*
			243-268 BL00790H 13.42 6.478e-27 455-481 BL00790M 8.74 8.683e-25 741-763 BL00790P 12.33 3.755e-24 830-855 BL00790F 15.90 5.200e-24 339-366 BL00790L 11.16 5.909e-21 721-741 BL00790A 19.74 1.964e-19 31-53
1396	BL00240	Receptor tyrosine kinase class III proteins.	BL00240F 17.74 9.500e-16 789-837 BL00240E 11.56 1.439e-15 736-774 BL00240G 28.45 8.793e-15 836-889
1396	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 3.647e-20 750-781 BL00107B 13.31 5.091e-13 818-834
1396	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109D 17.04 9.100e-22 819-842 PR00109E 14.41 7.429e-19 863-886 PR00109B 12.27 5.125e-18 750-769 PR00109A 15.00 2.895e-13 713-727 PR00109C 12.85 5.235e-12 800-811
1396	BL00239	Receptor tyrosine kinase class II proteins.	BL00239E 17.14 5.426e-27 790-840 BL00239B 25.15 3.000e-23 684-732 BL00239F 28.15 8.132e-21 844-889 BL00239D 16.81 2.143e-10 762-788 BL00239C 18.75 3.348e-10 737-760
1396	BL50001	Src homology 2 (SH2) domain proteins profile.	BL50001B 17.40 2.714e-11 747-768 BL50001D 11.00 7.300e-10 818-829 BL50001C 10.17 1.000e-09 799-810
1396	PR00014	FIBRONECTIN TYPE III REPEAT SIGNATURE	PR00014C 15.44 8.071e-13 490-509 PR00014B 14.77 3.400e-10 467-478 PR00014D 12.04 6.824e-10 508-523 PR00014A 8.22 3.455e-09 342-352
1401	BL00027	'Homeobox' domain proteins.	BL00027 26.43 2.000e-11 84-127
1403	BL01160	Kinesin light chain repeat proteins.	BL01160B 19.54 5.958e-09 387-441
1404	BL01113	CIq domain proteins.	BL01113B 18.26 2.500e-13 841-877
1406	BL01206	Amiloride-sensitive sodium channels proteins.	BL01206D 30.58 3.025e-28 363-412 BL01206G 21.72 6.063e-27 530-576 BL01206F 16.40 7.643e-15 485-506 BL01206E 20.72 5.650e-14 427-454 BL01206C 12.30 3.455e-12 333-352 BL01206B 13.56 1.205e-10 313-327
1408	BL01220	Phosphatidylethanolamine-binding protein family proteins.	BL01220B 16.65 1.000e-40 59-100 BL01220C 14.75 5.846e-34 100-128 BL01220A 22.62 3.400e-31 21-52
1409	BL00815	Alpha-isopropylmalate and homocitrate synthases proteins.	BL00815C 21.36 3.118e-09 786-815
1412	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 4.051e-09 1-16
1412	PR00806	VINCULIN SIGNATURE	PR00806B 4.28 9.640e-09 3-17
1418	PD00066	PROTEIN ZINC-FINGER METAL-BINDL	PD00066 13.92 5.200e-09 453-

SEQ ID NO:	Database entry ID	Description	Results*
			466
1418	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 7.882e-14 524-541 BL00028 16.07 8.269e-11 555-572 BL00028 16.07 2.543e-09 437-454 BL00028 16.07 4.600e-09 408-425 BL00028 16.07 6.657e-09 465-482
1418	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 3.160e-09 521-535 PR00048A 10.52 4.960e-09 434-448 PR00048A 10.52 6.760e-09 552-566 PR00048A 10.52 7.840e-09 462-476
1419	BL00022	EGF-like domain proteins.	BL00022A 7.48 5.000e-09 177-184 BL00022A 7.48 5.000e-09 241-248 BL00022A 7.48 8.000e-09 49-56
1419	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011D 14.03 5.696e-09 182-201 PR00011D 14.03 6.478e-09 86-105 PR00011D 14.03 9.087e-09 118-137
1419	DM01842	1 CELLULOSE-BINDING DOMAIN, BACTERIAL TYPE.	DM01842 11.31 9.922e-09 94-141
1421	PR00371	FLAVOPROTEIN PYRIDINE NUCLEOTIDE CYTOCHROME REDUCTASE SIGNATURE	PR00371D 14.55 4.536e-11 385-405
1421	PR00406	CYTOCHROME B5 REDUCTASE SIGNATURE	PR00406D 10.02 6.538e-10 385-405
1421	PR00409	PHTHALATE DIOXYGENASE REDUCTASE FAMILY SIGNATURE	PR00409F 12.70 2.484e-09 385-405
1421	PR00466	CYTOCHROME B-245 HEAVY CHAIN SIGNATURE	PR00466E 6.82 6.958e-17 386-404 PR00466C 10.17 8.244e-09 195-216
1422	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 4.462e-11 1087-1104
1422	PD00066	PROTEIN ZINC-FINGER METAL-BINDL	PD00066 13.92 6.478e-11 1075-1088
1422	BL00319	Amyloidogenic glycoprotein extracellular domain proteins.	BL00319C 17.12 4.375e-10 1154-1188
1422	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 4.522e-10 1084-1098 PR00048B 6.02 1.474e-09 1072-1082 PR00048A 10.52 6.760e-09 1056-1070
1423	PR00260	BACTERIAL CHEMOTAXIS SENSORY TRANSDUCER SIGNATURE	PR00260C 10.26 9.294e-09 146-167
1424	BL00845	CAP-Gly domain proteins.	BL00845 16.43 6.442e-21 405-430 BL00845 16.43 9.820e-19 203-228
1426	PD00066	PROTEIN ZINC-FINGER METAL-BINDL	PD00066 13.92 3.769e-15 369-382 PD00066 13.92 4.462e-15 285-298 PD00066 13.92 2.800e-14 257-270 PD00066 13.92 5.200e-14 313-326 PD00066 13.92 8.962e-10 341-354
1426	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 5.050e-13 269-286 BL00028 16.07 5.050e-13 297-314 BL00028 16.07 2.500e-10 325-342 BL00028 16.07 5.200e-10 353-370 BL00028

SEQ ID NO:	Database entry ID	Description	Results*
			16.07 7.000e-10 241-258 BL00028 16.07 9.700e-10 381-398
1426	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 8.500e-17 266-280 PR00048A 10.52 5.500e-14 294-308 PR00048A 10.52 4.706e-12 350-364 PR00048B 6.02 6.000e-12 310-320 PR00048B 6.02 6.538e-11 394-404 PR00048A 10.52 2.565e-10 238-252 PR00048B 6.02 2.688e-10 254-264 PR00048B 6.02 4.375e-10 338-348 PR00048A 10.52 5.304e-10 378-392 PR00048A 10.52 9.609e-10 322-336 PR00048B 6.02 5.263e-09 282-292 PR00048B 6.02 6.211e-09 366-376
1429	PR00671	INHIBIN BETA B CHAIN SIGNATURE	PR00671C 4.18 5.345e-09 9-29
1431	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803A 10.51 6.855e-09 215-236
1431	BL00226	Intermediate filaments proteins.	BL00226D 19.10 7.400e-09 390-437
1432	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803A 10.51 6.855e-09 251-272
1432	BL00226	Intermediate filaments proteins.	BL00226D 19.10 7.400e-09 426-473
1434	PR00545	RETINOIC ACID RECEPTOR SIGNATURE	PR00545A 5.35 9.430e-09 383-398
1436	BL01238	GDA1/CD39 family of nucleoside phosphatases proteins.	BL01238A 11.72 7.840e-16 76-91
1437	PD00930	PROTEIN GTPASE DOMAIN ACTIVATION.	PD00930B 33.72 2.800e-26 1256-1297 PD00930A 25.62 3.864e-13 1152-1178
1437	PF00620	GTPase-activator protein for Rho-like GTPases.	PF00620B 14.20 7.000e-12 1205-1222
1437	PR00683	SPECTRIN PLECKSTRIN HOMOLOGY DOMAIN SIGNATURE	PR00683B 16.62 2.603e-10 946-968 PR00683D 15.87 2.773e-09 1005-1024
1437	PR00543	OESTROGEN RECEPTOR SIGNATURE	PR00543H 10.86 7.573e-09 556-576
1437	PF00595	PDZ domain proteins (Also known as DHR or GLGF).	PF00595 13.40 7.600e-09 90-101
1437	BL00275	Shiga/ricin ribosomal inactivating toxins proteins signatu.	BL00275A 12.16 7.677e-09 1226-1240
1441	BL00223	Annexins repeat proteins domain proteins.	BL00223B 28.47 1.000e-40 140-190 BL00223C 24.79 1.000e-40 217-272 BL00223A 15.59 5.500e-32 21-55 BL00223A 15.59 4.783e-14 230-264 BL00223C 24.79 2.515e-10 8-63 BL00223A 15.59 6.250e-10 71-105
1441	PR00199	ANNEXIN TYPE III SIGNATURE	PR00199G 9.09 8.364e-21 239-265 PR00199F 16.19 5.636e-16 158-185 PR00199D 5.65 5.375e-14 25-47 PR00199B 6.86 1.574e-13 30-53 PR00199D 5.65 7.987e-

SEQ ID NO:	Database entry ID	Description	Results*
			13 234-256 PR00199H 12.62 5.339e-12 282-296 PR00199D 5.65 9.276e-10 75-97
1441	PR00200	ANNEXIN TYPE IV SIGNATURE	PR00200F 13.72 1.118e-35 158-185 PR00200G 9.43 1.000e-34 238-265 PR00200B 7.39 1.643e-29 30-53 PR00200H 13.68 1.766e-18 282-296 PR00200E 10.00 6.160e-16 75-97 PR00200I 10.00 2.111e-14 25-47 PR00200A 4.93 2.125e-14 5-16 PR00200C 8.76 1.500e-12 54-63 PR00200E 10.00 2.859e-11 234-256 PR00200G 9.43 5.294e-11 29-56 PR00200D 10.01 9.722e-10 70-87
1441	PR00197	ANNEXIN TYPE I SIGNATURE	PR00197F 9.03 5.250e-16 238-259 PR00197D 7.50 1.250e-15 25-47 PR00197E 11.89 8.463e-14 158-185 PR00197D 7.50 1.542e-12 234-256 PR00197D 7.50 5.451e-10 75-97 PR00197B 7.56 2.206e-09 30-53
1441	PR00198	ANNEXIN TYPE II SIGNATURE	PR00198G 8.09 7.943e-16 238-259 PR00198D 7.65 2.271e-13 234-256 PR00198D 7.65 9.894e-13 25-47 PR00198E 14.67 6.381e-11 158-185 PR00198H 12.05 1.462e-10 282-296 PR00198B 8.71 9.357e-10 30-53 PR00198D 7.65 4.845e-09 75-97
1441	PR00201	ANNEXIN TYPE V SIGNATURE	PR00201G 11.02 9.419e-26 238-265 PR00201A 6.05 4.770e-16 30-53 PR00201E 12.37 4.103e-15 158-185 PR00201H 12.04 4.375e-14 282-296 PR00201D 10.49 4.150e-10 75-97 PR00201G 11.02 8.402e-10 29-56 PR00201D 10.49 6.179e-09 25-47
1441	PR00301	70 KD HEAT SHOCK PROTEIN SIGNATURE	PR00301D 15.51 7.395e-09 38-59
1441	PR00196	ANNEXIN FAMILY SIGNATURE	PR00196D 21.86 3.032e-24 158-185 PR00196E 9.19 8.333e-23 238-259 PR00196A 11.16 9.100e-21 30-53 PR00196F 13.89 2.714e-15 266-282 PR00196C 10.36 5.167e-15 25-47 PR00196G 11.72 3.000e-14 282-296 PR00196C 10.36 7.344e-13 234-256 PR00196C 10.36 1.703e-12 75-97 PR00196G 11.72 9.217e-10 207-221 PR00196F 13.89 4.188e-09 107-123 PR00196A 11.16 7.840e-09 80-103
1441	PR00202	ANNEXIN TYPE VI SIGNATURE	PR00202G 8.01 4.833e-28 238-265 PR00202E 13.00 4.643e-16 158-185 PR00202D 5.58 9.604e-13 75-97 PR00202B 11.44 2.763e-11 29-53 PR00202H 9.20

SEQ ID NO:	Database entry ID	Description	Results*
			4.740e-11 282-296 PR00202D 5.58 1.908e-09 25-47 PR00202G 8.01 9.237e-09 29-56
1444	DM01513	CAMP-DEPENDENT PROTEIN KINASE REGULATORY CHAIN.	DM01513A 13.61 8.568e-14 15-56
1445	BL00603	Thymidine kinase cellular-type proteins.	BL00603C 30.02 1.000e-40 152-207 BL00603A 20.71 4.500e-33 63-96 BL00603D 10.53 5.091e-18 217-232 BL00603B 11.39 3.455e-15 132-147
1446	PD01922	PROTEIN PHOSPHODIESTERASE HYDROL.	PD01922B 21.83 7.328e-14 162-198
1447	BL00061	Short-chain dehydrogenases/reductases family proteins.	BL00061B 25.79 1.931e-13 99-137
1448	BL01160	Kinesin light chain repeat proteins.	BL01160B 19.54 5.958e-09 64-118
1449	PF00856	SET domain proteins.	PF00856A 26.14 8.579e-11 5-42
1449	PF00628	PHD-finger.	PF00628 15.84 5.500e-10 11-26
1452	BL00030	Eukaryotic RNA-binding region RNP-1 proteins.	BL00030B 7.03 3.400e-10 116-126
1454	PF00075	RNase H.	PF00075D 10.71 7.000e-11 517-528 PF00075C 11.58 9.786e-11 484-496 PF00075B 12.56 4.073e-10 449-460 PF00075A 14.44 2.143e-09 402-419
1454	PR00211	GLUTELIN SIGNATURE	PR00211B 0.86 4.417e-09 138-159
1456	BL00262	Insulin family proteins.	BL00262B 16.89 8.286e-17 68-88 BL00262A 12.48 4.600e-15 32-50
1456	PR00277	INSULIN B CHAIN SIGNATURE	PR00277A 14.82 2.421e-13 29-43 PR00277B 12.79 2.350e-11 43-56
1456	PR00276	INSULIN A CHAIN SIGNATURE	PR00276A 11.84 4.750e-13 69-79 PR00276B 8.02 7.828e-10 78-88
1457	PR00213	MYELIN P0 PROTEIN SIGNATURE	PR00213E 5.51 9.775e-12 264-289
1459	BL00856	Guanylate kinase proteins.	BL00856C 29.21 2.658e-26 539-587 BL00856B 9.61 2.946e-18 511-532
1459	PR00452	SH3 DOMAIN SIGNATURE	PR00452B 11.65 2.750e-09 369-385
1459	PD00289	PROTEIN SH3 DOMAIN REPEAT PRESNA.	PD00289 9.97 6.586e-09 298-312
1459	PF00595	PDZ domain proteins (Also known as DHR or GLGF).	PF00595 13.40 8.800e-09 295-306
1461	PR00475	HEXOKINASE FAMILY SIGNATURE	PR00475B 14.92 6.143e-26 186-212 PR00475E 16.08 2.742e-22 327-350 PR00475F 9.68 4.000e-20 407-430 PR00475A 14.06 3.118e-19 118-135 PR00475C 11.92 6.684e-19 239-256 PR00475G 9.08 1.692e-16 479-496 PR00475D 13.30 2.653e-13 262-277 PR00475G 9.08 2.650e-10 32-49
1461	BL00378	Hexokinases proteins.	BL00378C 16.14 1.000e-40 243-287 BL00378B 22.92 5.821e-40 313-359 BL00378D 14.23 3.647e-

SEQ ID NO:	Database entry ID	Description	Results*
			32 98-135 BL00378F 8.27 2.688e-17 481-496 BL00378D 10.94 1.474e-13 291-303 BL00378A 19.01 8.694e-11 59-87 BL00378F 8.27 3.714e-10 34-49
1464	PR00722	CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1) SIGNATURE	PR00722A 12.27 8.448e-14 56-72
1464	BL00021	Kringle domain proteins.	BL00021B 13.33 1.763e-13 55-73
1464	PR00839	V8 SERINE PROTEASE FAMILY SIGNATURE	PR00839B 11.20 4.945e-09 55-73
1464	BL01253	Type I fibronectin domain proteins.	BL01253E 16.01 6.381e-09 125-162
1464	BL00134	Serine proteases, trypsin family, histidine proteins.	BL00134A 11.96 3.813e-15 55-72 BL00134B 15.99 7.200e-10 186-210 BL00134C 13.45 9.206e-09 219-233
1466	BL00291	Prion protein.	BL00291A 4.49 9.379e-09 105-140
1467	PF00534	Glycosyl transferases group 1.	PF00534B 14.47 9.581e-12 398-422
1468	PF01105	emp24/gp25L/p24 family.	PF01105B 25.12 2.868e-25 126-178
1469	PF01105	emp24/gp25L/p24 family.	PF01105B 25.12 2.868e-25 151-203
1470	PR00305	14-3-3 PROTEIN ZETA SIGNATURE	PR00305A 9.33 9.500e-36 37-67 PR00305E 13.01 4.316e-32 177-204 PR00305D 16.34 3.647e-30 150-177 PR00305F 15.95 1.964e-26 204-234 PR00305C 8.68 3.182e-26 115-138 PR00305B 9.99 4.857e-24 84-109 PR00305F 15.95 8.975e-15 215-245
1470	BL00796	14-3-3 proteins.	BL00796C 17.44 1.000e-40 99-149 BL00796D 17.39 1.000e-40 150-196 BL00796B 10.67 7.000e-39 37-70 BL00796B 14.15 3.045e-33 198-234 BL00796A 10.52 4.656e-26 5-32 BL00796E 14.15 2.742e-11 209-245
1474	PF00642	Zinc finger C-x8-C-x5-C-x3-H type (and similar).	PF00642 11.59 7.796e-10 676-687 PF00642 11.59 7.055e-09 276-287
1475	PF00588	SpoU rRNA Methylase family.	PF00588B 17.18 8.200e-10 281-303
1476	BL01160	Kinesin light chain repeat proteins.	BL01160B 19.54 5.653e-09 791-845
1477	PR00326	GTP1/OBG GTP-BINDING PROTEIN FAMILY SIGNATURE	PR00326A 8.75 4.255e-14 364-385
1477	BL00306	Caseins alpha/beta proteins.	BL00306B 8.28 1.900e-09 557-568
1477	PR00318	ALPHA G-PROTEIN (TRANSDUCIN) SIGNATURE	PR00318E 7.23 5.320e-09 220-230
1479	BL01305	moaA / niIB / pggE family proteins.	BL01305D 14.97 7.279e-09 7-22
1480	PR00918	CALICIVIRUS NON-STRUCTURAL POLYPROTEIN FAMILY SIGNATURE	PR00918A 13.76 5.807e-09 458-479
1480	PR00674	LIGHT HARVESTING PROTEIN B CHAIN SIGNATURE	PR00674A 20.10 9.870e-09 133-154
1481	PR00171	SUGAR TRANSPORTER SIGNATURE	PR00171E 14.87 1.000e-08 73-86

SEQ ID NO:	Database entry ID	Description	Results*
1482	DM01418	352 FIBRILLAR COLLAGEN CARBOXYL-TERMINAL.	DM01418A 20.83 5.650e-23 101-149 DM01418B 22.51 8.500e-11 166-208 DM01418C 20.48 8.655e-10 236-278
1482	BL00291	Prion protein.	BL00291A 4.49 9.349e-10 23-58
1482	BL01113	C1q domain proteins.	BL01113A 17.99 6.114e-11 38-65 BL01113A 17.99 2.915e-10 35-62 BL01113A 17.99 6.538e-09 17-44 BL01113A 17.99 6.712e-09 29-56 BL01113A 17.99 8.442e-09 32-59
1483	DM01418	352 FIBRILLAR COLLAGEN CARBOXYL-TERMINAL.	DM01418A 20.83 5.650e-23 117-165 DM01418B 22.51 8.500e-11 182-224 DM01418C 20.48 8.655e-10 252-294
1483	BL00291	Prion protein.	BL00291A 4.49 9.349e-10 23-58
1483	BL01113	C1q domain proteins.	BL01113A 17.99 6.114e-11 38-65 BL01113A 17.99 2.915e-10 35-62 BL01113A 17.99 6.538e-09 17-44 BL01113A 17.99 6.712e-09 29-56 BL01113A 17.99 8.442e-09 32-59
1484	BL01052	Calponin family repeat proteins.	BL01052B 15.31 3.308e-11 30-56
1484	PR00888	SMOOTH MUSCLE PROTEIN/CALPONIN FAMILY SIGNATURE	PR00888C 12.27 2.141e-09 30-46
1486	BL00795	Involucrin proteins.	BL00795C 17.06 7.600e-09 239-284
1486	BL00415	Synapsins proteins.	BL00415N 4.29 9.409e-09 818-862
1490	BL01046	ATP-dependent serine proteases, lon family, serine active sit.	BL01046D 19.61 4.938e-35 452-493 BL01046C 17.03 9.581e-31 377-421 BL01046B 19.24 4.977e-29 331-377
1490	PR00830	ENDOPEPTIDASE LA (LON) SERINE PROTEASE (S16) SIGNATURE	PR00830D 8.08 2.552e-20 767-787 PR00830A 8.41 7.545e-18 375-395 PR00830E 13.94 8.500e-15 790-809 PR00830C 8.47 2.837e-13 737-757 PR00830B 14.73 7.429e-13 654-671
1490	BL01128	Shikimate kinase proteins.	BL01128A 18.84 8.027e-12 371-405
1490	PR00300	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT SIGNATURE	PR00300A 9.56 1.254e-10 371-390
1490	PR00819	CBXX/CFXQ SUPERFAMILY SIGNATURE	PR00819B 10.83 2.350e-10 370-386
1490	BL00674	AAA-protein family proteins.	BL00674B 4.46 8.071e-10 368-390
1490	PR00364	DISEASE RESISTANCE PROTEIN SIGNATURE	PR00364A 8.19 1.818e-09 370-386
1490	BL00113	Adenylate kinase proteins.	BL00113A 12.74 7.369e-09 372-389
1491	BL00824	Elongation factor I beta/beta'/delta chain proteins.	BL00824B 9.21 2.338e-09 150-170
1495	BL00615	C-type lectin domain proteins.	BL00615A 16.68 3.880e-11 47-65 BL00615B 12.25 2.286e-10 149-163
1498	PR00119	P-TYPE CATION-TRANSPORTING ATPASE SUPERFAMILY SIGNATURE	PR00119B 13.94 8.714e-12 35-50 PR00119E 8.48 7.716e-11 420-440

SEQ ID NO:	Database entry ID	Description	Results*
1498	PR00120	H ⁺ TRANSPORTING ATPASE (PROTON PUMP) SIGNATURE	PR00120C 9.90 7.037e-10 420-437
1498	BL00154	E1-E2 ATPases phosphorylation site proteins.	BL00154E 20.37 5.275e-19 263-304 BL00154F 8.23 6.175e-19 417-441 BL00154C 12.38 4.326e-13 31-50 BL00154D 12.57 5.935e-09 191-202
1499	BL00039	DEAD-box subfamily ATP-dependent helicases proteins.	BL00039D 21.67 3.455e-33 476-522 BL00039A 18.44 8.548e-23 145-184 BL00039C 15.63 8.500e-16 277-301 BL00039B 19.19 1.837e-12 191-217
1499	DM01537	kw SKI2W SKI2 NUCLEOLAR HELICASE.	DM01537B 21.63 8.990e-12 450-497
1499	PF00271	Helicases conserved C-terminal domain proteins.	PF00271 7.99 5.500e-10 507-515
1501	BL00115	Eukaryotic RNA polymerase II heptapeptide repeat proteins.	BL00115Z 3.12 9.669e-09 116-165
1502	PF00168	C2 domain proteins.	PF00168B 11.83 8.000e-10 38-49
1502	PR00360	C2 DOMAIN SIGNATURE	PR00360A 14.59 6.806e-10 43-56 PR00360B 13.61 2.227e-09 67-81 PR00360C 13.61 5.909e-09 223-237
1503	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 4.165e-13 780-835 PF00791B 28.49 6.767e-10 888-943 PF00791C 20.98 8.059e-09 794-833
1504	PF00023	Ank repeat proteins.	PF00023A 16.03 5.875e-10 437-453 PF00023A 16.03 7.000e-10 563-579 PF00023A 16.03 8.500e-10 248-264 PF00023A 16.03 9.250e-10 95-111 PF00023A 16.03 3.250e-09 596-612 PF00023A 16.03 3.893e-09 716-732 PF00023A 16.03 6.786e-09 62-78 PF00023A 16.03 9.036e-09 496-512
1504	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 2.957e-09 88-101 PD00078B 13.14 5.696e-09 556-569 PD00078B 13.14 9.217e-09 742-755
1504	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 5.024e-15 215-270 PF00791B 28.49 7.750e-14 62-117 PF00791B 28.49 9.795e-14 530-585 PF00791B 28.49 9.505e-13 683-738 PF00791B 28.49 7.253e-12 95-150 PF00791B 28.49 2.636e-11 716-771 PF00791C 20.98 5.696e-11 697-736 PF00791B 28.49 3.359e-10 404-459 PF00791B 28.49 5.369e-10 248-303 PF00791B 28.49 6.767e-10 563-618 PF00791C 20.98 8.052e-10 544-583 PF00791C 20.98 3.382e-09 229-268 PF00791B 28.49 7.275e-09 371-426 PF00791C 20.98 9.912e-09 385-424

SEQ ID NO:	Database entry ID	Description	Results*
1505	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479B 12.57 8.714e-09 143-159
1506	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479B 12.57 8.714e-09 167-183
1507	BL00518	Zinc finger, C3HC4 type (RING finger), proteins.	BL00518 12.23 5.286e-10 239-248
1507	BL01282	BIR repeat proteins.	BL01282B 30.49 1.900e-09 220-259
1507	BL00264	Neurohypophysial hormones proteins.	BL00264 8.98 4.884e-09 341-368
1510	BL00122	Carboxylesterases type-B serine proteins.	BL00122G 11.67 2.500e-15 15-26
1511	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 1.986e-11 340-353 PR00910A 2.51 1.986e-11 342-355 PR00910A 2.51 1.986e-11 344-357 PR00910A 2.51 9.778e-10 346-359 PR00910A 2.51 1.107e-09 338-351 PR00910A 2.51 3.464e-09 336-349
1511	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 4.508e-09 324-357
1512	DM01970	0 kw ZK632.12 YDR313C ENDOSOMAL III.	DM01970B 8.60 8.475e-15 175-188
1514	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 8.375e-10 149-164 PR00833H 2.30 2.846e-09 147-162
1514	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308A 5.90 9.630e-11 150-165 PR00308C 3.83 8.892e-10 104-114 PR00308C 3.83 8.892e-10 105-115 PR00308C 3.83 8.892e-10 151-161 PR00308C 3.83 8.892e-10 152-162 PR00308C 3.83 8.892e-10 153-163 PR00308C 3.83 8.892e-10 154-164 PR00308C 3.83 7.545e-09 103-113 PR00308C 3.83 7.896e-09 150-160 PR00308B 4.28 8.397e-09 150-162 PR00308A 5.90 9.047e-09 101-116
1514	PR00456	RIBOSOMAL PROTEIN P2 SIGNATURE	PR00456E 3.06 7.188e-10 144-159 PR00456E 3.06 1.684e-09 145-160 PR00456E 3.06 7.949e-09 97-112 PR00456E 3.06 9.430e-09 98-113
1515	PF00092	Troponin.	PF00092A 16.67 3.368e-09 448-483
1521	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 7.333e-09 322-336 PR00019B 11.36 9.280e-09 319-333
1522	BL00315	Dehydrins proteins.	BL00315A 9.35 7.197e-10 93-121
1524	PD00930	PROTEIN GTPASE DOMAIN ACTIVATION.	PD00930B 33.72 4.240e-16 235-276
1524	PR00234	HIV-1 MATRIX PROTEIN SIGNATURE	PR00234E 11.78 7.268e-09 361-375
1525	BL00183	Ubiquitin-conjugating enzymes proteins.	BL00183 28.97 8.338e-14 44-92
1527	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 6.850e-10 132-146 PR00019A 11.19 2.667e-09 135-149 PR00019B 11.36 9.640e-

SEQ ID NO:	Database entry ID	Description	Results*
			09 180-194 PR00019B 11.36 1.000e-08 277-291
1529	BL00625	Regulator of chromosome condensation (RCC1) proteins.	BL00625A 16.21 2.033e-16 567-596 BL00625B 17.69 4.205e-12 561-595 BL00625B 17.69 9.423e-11 93-127 BL00625B 17.69 1.444e-10 152-186 BL00625A 16.21 1.759e-10 99-128 BL00625A 16.21 2.739e-09 515-544 BL00625B 17.69 3.172e-09 43-77 BL00625A 16.21 4.170e-09 158-187
1529	PR00633	CHROMOSOME CONDENSATION REGULATOR RCC1 SIGNATURE	PR00633B 13.02 3.535e-09 561-575 PR00633A 9.32 6.260e-09 527-544 PR00633F 10.03 7.949e-09 528-543
1530	BL00414	Profilin proteins.	BL00414A 13.85 6.344e-13 2-16 BL00414E 15.46 6.283e-09 121-136
1530	PR00392	PROFILIN SIGNATURE	PR00392C 18.98 2.350e-12 41-62 PR00392A 14.10 4.176e-12 4-14 PR00392D 12.00 5.250e-11 63-78 PR00392F 17.40 7.955e-11 122-140 PR00392E 12.06 6.500e-09 109-123
1531	BL00414	Profilin proteins.	BL00414A 13.85 6.344e-13 2-16 BL00414E 15.46 6.283e-09 105-120
1531	PR00392	PROFILIN SIGNATURE	PR00392C 18.98 2.350e-12 41-62 PR00392A 14.10 4.176e-12 4-14 PR00392D 12.00 5.250e-11 63-78 PR00392F 17.40 7.955e-11 106-124 PR00392E 12.06 8.833e-09 93-107
1532	PD00301	PROTEIN REPEAT MUSCLE CALCIUM-BL.	PD00301A 10.24 8.200e-09 131-142
1533	DM01930	2 kw FINGER SMCX SMCY YDR096W.	DM01930F 14.16 1.310e-27 24-60
1534	BL00411	Kinesin motor domain proteins.	BL00411G 21.39 2.200e-39 77-119 BL00411H 15.66 8.800e-33 125-156 BL00411F 14.77 6.250e-18 33-58
1534	PR00380	KINESIN HEAVY CHAIN SIGNATURE	PR00380D 9.93 7.923e-26 126-148 PR00380C 13.18 1.000e-21 76-95 PR00380B 12.64 1.621e-16 42-60
1534	BL00893	mutT domain proteins.	BL00893 18.99 8.826e-09 176-201
1536	BL00600	Aminotransferases class-III pyridoxal-phosphate attachment si.	BL00600E 16.43 5.725e-15 164-193 BL00600G 12.43 7.000e-14 242-261 BL00600F 8.77 7.480e-11 207-220 BL00600D 8.71 1.750e-10 143-157
1537	BL00838	Interleukins -4 and -13 proteins.	BL00838A 12.35 8.696e-09 136-153
1537	PD01847	PHOTOSYSTEM II PROTEIN REACTION CENTRE I TRANS.	PD01847 9.59 8.946e-09 137-173
1539	PR00121	SODIUM/POTASSIUM-TRANSPORTING ATPASE SIGNATURE	PR00121D 16.72 3.012e-12 261-283

SEQ ID NO:	Database entry ID	Description	Results*
1539	BL00154	E1-E2 ATPases phosphorylation site proteins.	BL00154E 20.37 8.468e-16 532-573 BL00154C 12.38 3.520e-12 264-283
1539	BL01228	Hypothetical cof family proteins.	BL01228D 17.44 6.400e-11 660-685
1539	PR00119	P-TYPE CATION-TRANSPORTING ATPASE SUPERFAMILY SIGNATURE	PR00119B 13.94 3.333e-11 268-283 PR00119D 9.56 6.063e-10 548-559
1540	BL00289	Pentaxin family proteins.	BL00289A 30.36 9.031e-09 331-362
1542	BL01279	Protein-L-isoaspartate(D-aspartate) O-methyltransferase signa.	BL01279A 24.27 1.000e-11 67-115
1542	BL00422	Granins proteins.	BL00422C 16.18 7.176e-09 303-331
1545	BL00027	'Homeobox' domain proteins.	BL00027 26.43 4.462e-32 244-287
1545	PR00025	HOMEOTIC ANTENNAPEDIA PROTEIN SIGNATURE	PR00025B 11.94 3.143e-12 230-246
1545	PR00024	HOMEBOX SIGNATURE	PR00024C 7.49 3.500e-12 276-286 PR00024A 11.87 7.000e-12 251-263 PR00024B 11.27 1.409e-10 266-277
1545	PR00031	LAMBDA AND OTHER REPRESSOR HELIX-TURN-HELIX SIGNATURE	PR00031B 16.29 4.414e-10 267-284
1545	BL00032	'Homeobox' antennapedia-type protein.	BL00032B 10.83 1.675e-37 233-272 BL00032C 11.28 4.429e-21 272-290 BL00032A 18.38 5.750e-10 193-216
1546	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 6.400e-16 245-258 PD00066 13.92 8.615e-15 329-342 PD00066 13.92 6.000e-13 301-314 PD00066 13.92 4.857e-12 217-230 PD00066 13.92 1.346e-10 273-285 PD00066 13.92 8.200e-09 357-370
1546	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 5.950e-13 313-330 BL00028 16.07 7.261e-12 229-246 BL00028 16.07 3.077e-11 16-33 BL00028 16.07 3.769e-11 283-302 BL00028 16.07 9.308e-11 341-358 BL00028 16.07 3.100e-10 397-414 BL00028 16.07 5.800e-10 201-218 BL00028 16.07 6.400e-10 369-386 BL00028 16.07 7.600e-10 257-274 BL00028 16.07 8.800e-10 72-89 BL00028 16.07 9.229e-09 101-118
1546	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 3.118e-12 310-324 PR00048B 6.02 5.000e-12 326-336 PR00048A 10.52 6.294e-12 13-27 PR00048B 6.02 1.692e-11 242-252 PR00048A 10.52 3.842e-11 338-352 PR00048A 10.52 5.263e-11 366-380 PR00048A 10.52 8.579e-11 226-240 PR00048A 10.52 8.579e-11

SEQ ID NO:	Database entry ID	Description	Results*
			254-268 PR00048A 10.52 3.348e-10 394-408 PR00048A 10.52 4.913e-10 282-296 PR00048B 6.02 7.188e-10 298-308 PR00048B 6.02 9.053e-09 57-67 PR00048A 10.52 9.640e-09 98-112
1547	BL00585	Ribosomal protein S5 proteins.	BL00585B 18.78 6.143e-18 303-340 BL00585A 28.43 4.286e-16 220-272
1548	PR00482	OMPTIN SERINE PROTEASE SIGNATURE	PR00482C 11.02 7.968e-09 816-842
1549	PR00500	POLYCYSTIC KIDNEY DISEASE PROTEIN SIGNATURE	PR00500B 7.74 7.359e-10 56-77
1551	PR00917	SMALL ROUND STRUCTURED VIRUS (C37) CYSTEINE PROTEASE FAMILY SIGNATURE	PR00917G 10.59 8.990e-09 812-830
1553	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 9.486e-09 109-126
1555	PF00638	RanBP1 domain proteins.	PF00638 11.91 4.600e-18 67-82
1555	DM01269	303 kw ACTIVATING RAN GTPASE ISOZYME.	DM01269A 23.35 1.600e-20 68-96 DM01269B 11.71 3.323e-09 138-148
1556	BL00406	Actins proteins.	BL00406E 8.44 8.541e-28 323-373 BL00406B 5.47 1.375e-27 82-137 BL00406D 12.58 3.160e-26 266-321 BL00406C 6.75 6.943e-25 141-196 BL00406A 9.95 2.575e-20 7-42
1556	PR00190	ACTIN SIGNATURE	PR00190F 7.80 3.647e-13 139-159 PR00190C 11.49 2.029e-12 60-83 PR00190G 12.62 2.050e-09 233-250
1558	BL00048	Protamine P1 proteins.	BL00048 6.39 3.700e-09 153-180
1558	DM01206	CORONAVIRUS NUCLEOCAPSID PROTEIN.	DM01206B 10.69 9.328e-11 157-177 DM01206B 10.69 1.247e-10 236-256 DM01206B 10.69 7.781e-10 188-208 DM01206B 10.69 6.582e-09 234-254
1559	PR00315	GTP-BINDING ELONGATION FACTOR SIGNATURE	PR00315A 11.81 5.688e-10 126-140
1559	PR00326	GTP/DBG GTP-BINDING PROTEIN FAMILY SIGNATURE	PR00326A 8.75 1.000e-09 127-148
1559	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 6.431e-09 125-147
1559	PR00755	AFLATOXIN BIOSYNTHESIS REGULATORY PROTEIN SIGNATURE	PR00755F 10.99 9.722e-09 30-52
1563	PF00780	Domain found in NIK1-like kinases, mouse citron and yeast ROM.	PF00780B 23.03 9.908e-09 14-57
1567	BL00162	Eukaryotic-type carbonic anhydrases proteins.	BL00162C 17.78 1.000e-40 88-125 BL00162E 14.93 7.231e-39 171-204 BL00162F 22.68 5.050e-31 208-242 BL00162A 22.92 8.714e-30 16-47 BL00162D 15.06 7.158e-24 126-151 BL00162B 21.43 1.375e-19 51-74
1568	PR00457	ANIMAL HAEM PEROXIDASE SIGNATURE	PR00457E 20.67 1.621e-24 414-441 PR00457D 16.81 8.258e-21

SEQ ID NO:	Database entry ID	Description	Results*
			389-410 PR00457B 13.29 3.455e-18 223-239 PR00457G 17.45 7.000e-18 595-616 PR00457C 19.25 4.414e-16 371-390 PR00457H 15.90 8.630e-14 666-681 PR00457A 15.80 5.645e-12 169-181 PR00457F 13.69 8.875e-11 467-478
1569	BL01304	ubiH/COQ6 monooxygenase family proteins.	BL01304A 8.05 3.571e-11 50-64
1569	PR00368	FAD-DEPENDENT PYRIDINE NUCLEOTIDE REDUCTASE SIGNATURE	PR00368A 17.76 3.769e-10 50-73
1569	PR00757	FLAVIN-CONTAINING AMINE OXIDASE SIGNATURE	PR00757A 6.64 5.552e-10 50-70
1569	BL00623	GMC oxidoreductases proteins.	BL00623A 12.60 2.929e-09 50-69
1569	PR00420	AROMATIC-RING HYDROXYLASE (FLAVOPROTEIN MONOOXYGENASE) SIGNATURE	PR00420A 14.78 6.455e-09 50-73
1569	BL00064	L-lactate dehydrogenase proteins.	BL00064A 21.16 7.203e-09 50-88
1569	PR00370	FLAVIN-CONTAINING MONOOXYGENASE (FMO) SIGNATURE	PR00370A 3.35 9.772e-09 50-66
1571	BL00019	Actinin-type actin-binding domain proteins.	BL00019D 15.33 3.880e-17 145-175
1573	BL00893	mutT domain proteins.	BL00893 18.99 5.500e-16 127-152
1573	PR00502	MUTT DOMAIN SIGNATURE	PR00502B 15.92 4.600e-13 138-154 PR00502A 15.06 2.636e-09 124-139
1574	PF00632	HECT-domain (ubiquitin-transferase).	PF00632B 18.45 7.000e-16 488-516 PF00632C 20.66 7.851e-14 533-565
1576	PR00239	MOLLUSCAN RHODOPSIN C-TERMINAL TAIL SIGNATURE	PR00239E 1.58 9.566e-10 292-304
1576	BL00115	Eukaryotic RNA polymerase II heptapeptide repeat proteins.	BL00115Z 3.12 5.632e-09 243-292
1576	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 6.118e-11 296-329 DM00215 19.43 9.647e-11 327-360 DM00215 19.43 8.232e-10 322-355 DM00215 19.43 2.068e-09 291-324 DM00215 19.43 2.983e-09 265-298 DM00215 19.43 4.356e-09 292-325 DM00215 19.43 7.712e-09 275-308 DM00215 19.43 8.017e-09 266-299 DM00215 19.43 8.475e-09 271-304 DM00215 19.43 8.780e-09 286-319
1582	BL01280	Glucose inhibited division protein A family proteins.	BL01280A 15.97 6.727e-36 69-110 BL01280B 23.56 8.105e-27 128-180
1582	BL00076	Pyridine nucleotide-disulphide oxidoreductases class-I.	BL00076A 18.83 6.745e-12 68-98
1582	BL00836	Alanine dehydrogenase & pyridine nucleotide transhydrogenase.	BL00836D 22.30 9.576e-12 69-106
1582	BL00504	Fumarate reductase / succinate dehydrogenase FAD-binding site proteins.	BL00504A 10.76 3.870e-11 69-91
1582	BL00977	FAD-dependent glycerol-3-phosphate dehydrogenase proteins.	BL00977A 20.76 8.583e-11 69-121

SEQ ID NO:	Database entry ID	Description	Results*
1582	PR00411	PYRIDINE NUCLEOTIDE DISULPHIDE REDUCTASE CLASS-I SIGNATURE	PR00411A 15.95 1.000e-10 69-92
1582	BL00982	Bacterial-type phytoene dehydrogenase proteins.	BL00982A 18.41 2.151e-10 71-103
1582	PR00368	FAD-DEPENDENT PYRIDINE NUCLEOTIDE REDUCTASE SIGNATURE	PR00368A 17.76 8.846e-13 69-92 PR00368C 15.74 5.263e-10 69-95
1582	PR00419	ADRENODOXIN REDUCTASE FAMILY SIGNATURE	PR00419A 14.89 3.571e-09 69-92
1582	PR00757	FLAVIN-CONTAINING AMINE OXIDASE SIGNATURE	PR00757A 6.64 6.226e-09 69-89
1582	PR00469	PYRIDINE NUCLEOTIDE DISULPHIDE REDUCTASE CLASS-II SIGNATURE	PR00469A 15.46 1.851e-10 69-92 PR00469F 16.51 8.063e-09 65-90
1582	BL00623	GMC oxidoreductases proteins.	BL00623A 12.60 8.586e-09 69-88
1586	PR00413	HALOACID DEHALOGENASE/EPOXIDE HYDROLASE FAMILY SIGNATURE	PR00413E 15.78 6.714e-09 70-87
1587	PD01861	PROTEIN NUCLEAR RIBONUCLEOPROTEIN SMALL MRNA RNA.	PD01861A 14.06 6.318e-10 60-84
1588	PD00289	PROTEIN SH3 DOMAIN REPEAT PRESYN.	PD00289 9.97 6.586e-09 46-60
1588	PF00595	PDZ domain proteins (Also known as DHR or GLGF).	PF00595 13.40 9.400e-09 43-54
1591	BL00914	Syntaxin / epimorphin family proteins.	BL00914 24.91 1.250e-29 184-234
1592	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010C 11.16 6.667e-11 363-374
1592	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 6.538e-16 358-374 BL01187A 9.98 3.250e-09 278-290
1592	BL00022	EGF-like domain proteins.	BL00022B 7.54 8.200e-09 367-374
1593	PR00625	DNAJ PROTEIN FAMILY SIGNATURE	PR00625A 12.84 4.600e-20 14-34 PR00625B 13.48 8.759e-20 46-67
1593	BL00636	Nt-dnaJ domain proteins.	BL00636A 8.07 4.176e-18 18-35 BL00636B 15.11 1.000e-15 46-67
1594	PD02448	TRANSCRIPTION PROTEIN DNA-BINDIN.	PD02448A 9.37 3.854e-09 351-390
1598	PD02448	TRANSCRIPTION PROTEIN DNA-BINDIN.	PD02448A 9.37 1.511e-20 50-89 PD02448B 10.17 8.071e-19 89-137
1602	PR00403	WW DOMAIN SIGNATURE	PR00403B 12.19 9.816e-11 144-159 PR00403B 12.19 8.167e-10 103-118
1602	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 6.644e-09 23-56
1602	BL01159	WW/rsp5/WWP domain proteins.	BL01159 13.85 5.224e-10 144-159 BL01159 13.85 6.891e-09 103-118
1602	PR00571	ENDOTHELIN-B RECEPTOR SIGNATURE	PR00571G 5.36 7.750e-09 107-126
1603	PR00403	WW DOMAIN SIGNATURE	PR00403B 12.19 9.816e-11 107-122
1603	BL01159	WW/rsp5/WWP domain proteins.	BL01159 13.85 5.224e-10 107-122
1603	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 6.644e-09 23-56
1605	PR00929	AT-HOOK-LIKE DOMAIN SIGNATURE	PR00929B 4.38 4.600e-10 358-370
1605	DM00303	6 LEA 11-MER REPEAT REPEAT.	DM00303A 13.20 7.708e-10 306-

SEQ ID NO:	Database entry ID	Description	Results*
			356 DM00303A 13.20 2.912e-09 304-354 DM00303A 13.20 7.212e-09 300-350 DM00303A 13.20 7.212e-09 311-361
1605	BL00354	HMG-I and HMG-Y DNA-binding domain proteins (Ahook).	BL00354B 3.16 7.722e-09 357-370
1606	PD02379	AMINOTRANSFERASE BIOSYNTHESIS PHOSPHOSERINE SER.	PD02379E 11.43 1.000e-40 194-236 PD02379F 18.62 6.029e-35 245-284 PD02379H 16.03 5.235e-33 352-385 PD02379B 12.05 3.613e-31 80-113 PD02379A 15.57 2.800e-25 29-60 PD02379C 13.34 3.700e-21 119-139 PD02379D 11.83 9.419e-16 168-181 PD02379G 10.62 2.537e-14 313-328
1606	PR00800	AROMATIC-L-AMINO-ACID DECARBOXYLASE SIGNATURE	PR00800G 11.29 1.889e-09 97-118
1607	PD02379	AMINOTRANSFERASE BIOSYNTHESIS PHOSPHOSERINE SER.	PD02379E 11.43 1.000e-40 194-236 PD02379F 18.62 6.029e-35 245-284 PD02379B 12.05 3.613e-31 80-113 PD02379A 15.57 2.800e-25 29-60 PD02379H 16.03 7.864e-23 306-339 PD02379C 13.34 3.700e-21 119-139 PD02379D 11.83 9.419e-16 168-181
1607	PR00800	AROMATIC-L-AMINO-ACID DECARBOXYLASE SIGNATURE	PR00800G 11.29 1.889e-09 97-118
1610	PR00874	FUNGI-IV METALLOTHIONEIN SIGNATURE	PR00874C 4.37 6.625e-09 33-48
1614	BL00035	'POU' domain proteins.	BL00035B 14.46 6.236e-09 683-704
1616	PF00168	C2 domain proteins.	PF00168C 27.49 8.412e-13 634-660
1616	PR00360	C2 DOMAIN SIGNATURE	PR00360B 13.61 8.105e-10 651-665
1617	PF00168	C2 domain proteins.	PF00168C 27.49 8.412e-13 115-141
1617	PR00360	C2 DOMAIN SIGNATURE	PR00360B 13.61 8.105e-10 132-146
1619	PR00566	DOPAMINE 1B RECEPTOR SIGNATURE	PR00566E 13.44 5.255e-18 466-483 PR00566A 9.32 3.000e-17 200-214 PR00566D 9.35 1.600e-12 446-455 PR00566C 11.44 2.184e-12 401-412 PR00566B 8.20 3.053e-11 341-351
1619	PR00242	DOPAMINE RECEPTOR SIGNATURE	PR00242E 13.29 1.000e-12 424-439 PR00242B 11.77 8.650e-11 257-267
1619	BL00237	G-protein coupled receptors proteins.	BL00237C 13.19 6.786e-20 364-391 BL00237A 27.68 9.710e-15 266-306 BL00237B 5.28 5.263e-10 309-321
1619	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237F 13.57 5.800e-19 369-394 PR00237B 13.50 6.250e-19 236-258 PR00237E 13.03 9.500e-15 301-325 PR00237C 15.69

SEQ ID NO:	Database entry ID	Description	Results*
			3.925e-09 280-303 PR00237A 11.48 7.387e-09 202-227
1620	PR00169	POTASSIUM CHANNEL SIGNATURE	PR00169A 16.77 7.851e-11 46-66
1621	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 7.529e-11 183-216
1621	PF00685	Sulfotransferase proteins.	PF00685C 26.03 5.100e-09 118-164
1621	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 8.322e-09 198-213
1622	BL00951	ER lumen protein retaining receptor proteins.	BL00951B 14.23 1.670e-09 43-74
1623	BL00292	Cyclins proteins.	BL00292B 20.31 3.925e-11 120-151
1624	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 3.160e-09 111-125
1624	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 3.739e-12 114-131 BL00028 16.07 3.571e-09 145-162
1625	BL01226	Hydroxymethylglutaryl-coenzyme A synthase proteins.	BL012261 25.06 8.560e-09 256-304
1629	BL00030	Eukaryotic RNA-binding region RNP-1 proteins.	BL00030A 14.39 1.563e-12 72-91 BL00030A 14.39 2.125e-12 156-175
1637	BL00740	MAM domain proteins.	BL00740B 19.76 3.813e-09 637-658
1637	PR00597	GELSOLIN FAMILY SIGNATURE	PR00597G 8.55 6.586e-27 637-660 PR00597A 12.96 5.846e-26 326-348 PR00597E 13.46 2.000e-22 523-544 PR00597F 16.29 9.526e-22 582-602 PR00597D 12.77 1.000e-20 469-490 PR00597B 9.78 2.500e-20 415-432 PR00597C 14.19 6.192e-20 436-455 PR00597H 15.32 7.577e-19 666-686 PR00597D 12.77 3.392e-10 94-115 PR00597B 9.78 9.455e-10 36-53 PR00597C 14.19 7.875e-09 61-80 PR00597A 12.96 8.027e-09 689-711
1641	BL00027	'Homeobox' domain proteins.	BL00027 26.43 7.000e-11 93-136
1641	PR00887	STRUCTURE-SPECIFIC RECOGNITION PROTEIN SIGNATURE	PR00887D 15.12 8.909e-09 337-351
1646	PR00259	TRANSMEMBRANE FOUR FAMILY SIGNATURE	PR00259A 9.27 3.308e-18 19-43 PR00259C 16.40 9.800e-18 88-117 PR00259D 13.50 2.756e-15 238-265
1646	BL00421	Transmembrane 4 family proteins.	BL00421A 11.79 5.263e-14 15-34 BL00421E 20.97 4.632e-13 235-265
1651	PR00669	INHIBIN ALPHA CHAIN SIGNATURE	PR00669F 5.57 9.899e-09 223-241
1652	DM01292	ESICULAR LUMEN DOMAIN.	DM01292L 12.54 9.505e-09 240-265
1653	PR00128	COLIPASE SIGNATURE	PR00128D 9.77 6.250e-25 47-66 PR00128C 9.28 5.299e-20 24-47
1653	BL00121	Colipase proteins.	BL00121B 9.96 3.160e-33 15-64 BL00121A 14.36 2.107e-09 16-56
1656	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 9.929e-10 384-

SEQ ID NO:	Database entry ID	Description	Results ^a
			399
1658	BL01118	Translation initiation factor SUI1 proteins.	BL01118B 26.75 8.579e-26 94-132 BL01118A 12.46 4.000e-13 77-92
1659	BL00811	Oleocins proteins.	BL00811A 8.26 3.310e-09 120-158
1660	BL00674	AAA-protein family proteins.	BL00674B 4.46 9.182e-11 184-206
1660	PR00830	ENDOPEPTIDASE LA (LON) SERINE PROTEASE (S16) SIGNATURE	PR00830A 8.41 8.544e-10 191-211
1660	PR00300	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT SIGNATURE	PR00300A 9.56 9.416e-09 187-206
1660	PR00051	BACTERIAL CHROMOSOMAL REPLICATION INITIATOR (DNAA) SIGNATURE	PR00051A 10.68 9.899e-09 184-205
1661	DM01871	kw SSR LIGASE CYCLO FORMYL TETRAHYDROFOLATE.	DM01871C 20.79 9.836e-10 270-296
1663	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 7.712e-09 95-110
1665	BL01181	Ribosomal protein S21 proteins.	BL01181 15.43 2.500e-10 13-49
1666	PR00259	TRANSMEMBRANE FOUR FAMILY SIGNATURE	PR00259C 16.40 6.824e-16 88-117 PR00259A 9.27 3.423e-14 24-48 PR00259D 13.50 1.574e-13 238-265 PR00259B 14.81 8.714e-13 61-88
1666	BL00421	Transmembrane 4 family proteins.	BL00421B 17.62 4.600e-19 67-106 BL00421E 20.97 6.211e-13 235-265 BL00421A 11.79 5.600e-12 20-39
1668	PR00496	NAPIN SIGNATURE	PR00496A 6.68 6.276e-09 21-43
1671	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 2.543e-11 670-697 PR00237A 11.48 3.000e-10 424-449
1671	PR00373	GLYCOPROTEIN HORMONE RECEPTOR SIGNATURE	PR00373D 11.16 2.403e-09 503-518
1671	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 6.600e-10 496-536 BL00237D 11.23 4.545e-09 680-697
1671	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 7.429e-09 400-413
1671	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 6.000e-11 94-108 PR00019A 11.19 7.300e-11 215-229 PR00019B 11.36 6.850e-10 46-60 PR00019A 11.19 8.043e-10 285-299 PR00019B 11.36 5.320e-09 212-226 PR00019B 11.36 9.640e-09 70-84
1672	BL00972	Ubiquitin carboxyl-terminal hydrolases family 2 proteins.	BL00972A 11.93 7.500e-20 36-54 BL00972D 22.55 6.806e-16 296-321 BL00972B 9.45 1.000e-13 116-126 BL00972E 20.72 8.773e-12 321-343
1673	PF00646	F-box domain proteins.	PF00646A 14.37 6.906e-09 92-106
1675	BL00933	FGGY family of carbohydrate kinases proteins.	BL00933D 24.01 7.545e-15 212-249 BL00933B 15.94 2.200e-09 54-65 BL00933E 13.80 3.543e-09 439-455 BL00933A 17.50

SEQ ID NO:	Database entry ID	Description	Results*
			4.857e-09 20-44
1676	PR00876	NEMATODE METALLOTHIONEIN SIGNATURE	PR00876B 7.66 1.887e-10 137-151
1676	BL00026	Chitin recognition or binding domain proteins.	BL00026 12.95 2.776e-09 55-76
1676	BL01208	VWFC domain proteins.	BL01208B 15.83 2.946e-09 172-187
1676	BL01282	BIR repeat proteins.	BL01282B 30.49 4.471e-09 130-169
1676	PR00875	MOLLUSC METALLOTHIONEIN SIGNATURE	PR00875B 5.24 4.649e-09 137-145
1676	BL00956	Fungal hydrophobins proteins.	BL00956B 8.29 4.682e-09 153-165
1676	PD00866	GLYCOPROTEIN PROTEIN SPIKE E2 PRECURSOR PEPLIMER.	PD00866L 3.73 6.564e-10 1-11 PD00866L 3.73 1.443e-09 33-43 PD00866L 3.73 2.918e-09 12-22 PD00866L 3.73 2.918e-09 19-29 PD00866L 3.73 4.836e-09 149-159
1676	PD02283	PROTEIN SPORULATION REPEAT PRECU.	PD02283C 17.54 6.288e-09 141-169
1676	DM01724	kw ALLERGEN POLLEN CIM1 HOL-LI.	DM01724 8.14 7.465e-10 17-37 DM01724 8.14 4.434e-09 19-39 DM01724 8.14 6.684e-09 10-30
1676	PR00858	CRUSTACEAN METALLOTHIONEIN SIGNATURE	PR00858B 5.93 5.883e-09 155-174 PR00858B 5.93 8.085e-09 136-155
1676	PR00874	FUNGI-IV METALLOTHIONEIN SIGNATURE	PR00874C 4.37 9.739e-10 125-140 PR00874C 4.37 9.000e-09 135-150
1676	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243I 31.77 3.779e-10 38-81 BL00243I 31.77 4.309e-10 68-111 BL00243I 31.77 5.235e-10 58-101 BL00243I 31.77 7.353e-10 98-141 BL00243I 31.77 1.000e-09 78-121 BL00243I 31.77 1.000e-09 88-131 BL00243I 31.77 1.380e-09 121-164 BL00243I 31.77 2.648e-09 119-162 BL00243I 31.77 3.662e-09 61-104 BL00243I 31.77 4.296e-09 131-174 BL00243I 31.77 4.676e-09 48-91 BL00243I 31.77 6.704e-09 109-152 BL00243I 31.77 7.845e-09 25-68 BL00243I 31.77 9.366e-09 134-177
1676	BL00203	Vertebrate metallothioneins proteins.	BL00203 13.94 3.885e-16 128-174 BL00203 13.94 8.607e-13 123-169 BL00203 13.94 2.780e-11 153-199 BL00203 13.94 3.571e-11 148-194 BL00203 13.94 4.363e-11 113-159 BL00203 13.94 5.451e-11 139-185 BL00203 13.94 6.934e-11 144-190 BL00203 13.94 9.209e-11 131-177 BL00203 13.94 2.436e-10 35-81 BL00203 13.94 4.255e-10 133-179 BL00203

SEQ ID NO:	Database entry ID	Description	Results*
			13.94 6.745e-10 32-78 BL00203 13.94 7.032e-10 154-200 BL00203 13.94 2.929e-09 34-80 BL00203 13.94 3.388e-09 149-195 BL00203 13.94 3.571e-09 136-182 BL00203 13.94 5.224e-09 127-173 BL00203 13.94 5.776e-09 43-89 BL00203 13.94 6.878e-09 140-186 BL00203 13.94 7.796e-09 45-91 BL00203 13.94 9.541e-09 42-88
1679	PD01976	KINASE DEHYDROGENASE TRANSFERASE.	PD01976A 8.95 1.493e-09 83-96
1680	BL00623	GMC oxidoreductases proteins.	BL00623A 12.60 9.859e-10 12-31
1680	PR00419	ADRENODOXIN REDUCTASE FAMILY SIGNATURE	PR00419A 14.89 4.729e-09 12-35
1680	PR00368	FAD-DEPENDENT PYRIDINE NUCLEOTIDE REDUCTASE SIGNATURE	PR00368A 17.76 9.357e-09 12-35
1683	BL01172	Ribosomal protein L44e proteins.	BL01172B 14.10 8.909e-38 15-57 BL01172C 16.78 7.188e-31 63-102
1685	DM01724	kw ALLERGEN POLLEN CIM1 HOL-LI.	DM01724 8.14 5.909e-11 11-31 DM01724 8.14 6.591e-11 41-61 DM01724 8.14 6.831e-10 39-59 DM01724 8.14 8.697e-09 55-75
1686	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320A 16.74 8.463e-09 73-88
1690	PD02269	CYTIDINE DEAMINASE HYDROLASE ZINC AMINOHY.	PD02269C 16.36 7.882e-17 79-92 PD02269A 10.06 1.000e-15 29-41 PD02269D 11.98 5.000e-14 110-125
1691	BL00790	Receptor tyrosine kinase class V proteins.	BL00790D 12.41 8.297e-09 429-454
1692	BL00750	Chaperonins TCP-1 proteins.	BL00750B 16.17 2.000e-39 69-119 BL00750A 20.07 8.286e-36 25-68 BL00750C 25.65 8.579e-23 152-184
1692	PR00304	TAILLESS COMPLEX POLYPEPTIDE 1 (CHAPERONE) SIGNATURE	PR00304C 8.69 1.250e-18 86-106 PR00304B 11.60 2.059e-17 56-75 PR00304A 9.20 3.605e-15 34-51
1692	PR00298	60 KD CHAPERONIN SIGNATURE	PR00298B 13.59 7.353e-11 88-116
1692	BL00296	Chaperonins cpn60 proteins.	BL00296B 15.98 4.115e-13 76-130 BL00296A 17.20 5.648e-10 12-66
1694	BL00415	Synapsins proteins.	BL00415N 4.29 4.710e-10 225-269
1694	PD01234	PROTEIN NUCLEAR BROMODOMAIN TRANS.	PD01234B 15.53 5.875e-10 243-261
1694	BL00795	Involucrin proteins.	BL00795C 17.06 7.698e-10 213-258
1694	PR00208	GLIADIN AND LMW GLUTENIN SUPERFAMILY SIGNATURE	PR00208A 12.59 9.384e-09 247-265
1694	DM00406	GLIADIN.	DM00406 7.73 9.800e-09 245-258
1696	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 1.000e-09 212-229 BL00028 16.07 6.143e-09 365-382

SEQ ID NO:	Database entry ID	Description	Results*
1700	BL00030	Eukaryotic RNA-binding region RNP-1 proteins.	BL00030A 14.39 6.143e-13 332-351
1700	PD02784	PROTEIN NUCLEAR RIBONUCLEOPROTEIN.	PD02784B 26.46 6.943e-09 442-485
1701	PF00023	Ank repeat proteins.	PF00023A 16.03 8.500e-10 283-299 PF00023A 16.03 9.625e-10 347-363 PF00023A 16.03 1.321e-09 184-200 PF00023A 16.03 1.643e-09 150-166
1701	BL00906	Uroporphyrinogen decarboxylase proteins.	BL00906D 24.33 7.750e-09 212-256
1701	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 8.159e-14 117-172 PF00791B 28.49 8.319e-13 217-272 PF00791B 28.49 3.179e-12 184-239 PF00791B 28.49 5.168e-12 347-402 PF00791B 28.49 5.727e-11 250-305 PF00791B 28.49 2.817e-09 17-72 PF00791B 28.49 8.514e-09 84-139 PF00791C 20.98 1.000e-08 98-137
1702	BL00107	Protein kinases ATP-binding region proteins.	BL00107B 13.31 1.643e-10 202-218
1702	PF00992	Troponin.	PF00992A 16.67 9.526e-09 749-784
1708	PR00671	INHIBIN BETA B CHAIN SIGNATURE	PR00671C 4.18 8.966e-09 212-232
1709	PR00678	P13 KINASE P85 REGULATORY SUBUNIT SIGNATURE	PR00678H 9.13 7.805e-12 292-315
1710	PR00412	EPOXIDE HYDROLASE SIGNATURE	PR00412C 11.30 2.421e-12 169-183 PR00412A 13.23 7.947e-12 104-123 PR00412B 12.59 7.429e-10 123-139
1711	PR00217	43 KD POSTSYNAPTIC PROTEIN SIGNATURE	PR00217C 10.91 7.247e-10 293-309
1712	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320A 16.74 9.122e-09 277-292 PR00320A 16.74 9.780e-09 233-248 PR00320C 13.01 1.000e-08 233-248
1713	BL01230	RNA methyltransferase trmA family proteins.	BL01230E 15.79 2.918e-11 487-503
1719	BL00039	DEAD-box subfamily ATP-dependent helicases proteins.	BL00039D 21.67 2.957e-09 434-480
1719	DM01537	kw SKI2W SKI2 NUCLEOLAR HELICASE.	DM01537B 21.63 7.830e-09 408-455
1721	PR00453	VON WILLEBRAND FACTOR TYPE A DOMAIN SIGNATURE	PR00453A 12.79 2.957e-10 33-51
1721	PR00527	GASTRIN RECEPTOR SIGNATURE	PR00527I 5.36 6.559e-09 419-439
1721	PR00014	FIBRONECTIN TYPE III REPEAT SIGNATURE	PR00014C 15.44 6.870e-09 381-400
1721	PR00477	PHOSPHOGLYCERATE KINASE FAMILY SIGNATURE	PR00477I 8.53 1.000e-08 168-186
1725	PR00493	BREAST CANCER TYPE I SUSCEPTIBILITY PROTEIN SIGNATURE	PR00493G 7.57 3.711e-14 693-714
1726	BL00443	Glutamine amidotransferases class-II proteins.	BL00443F 16.68 8.714e-09 85-101
1728	DM01206	CORONAVIRUS NUCLEOCAPSID	DM01206B 10.69 7.288e-10 167-

SEQ ID NO:	Database entry ID	Description	Results*
		PROTEIN.	187
1728	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 3.411e-10 331-364 DM00215 19.43 7.107e-10 336-369 DM00215 19.43 9.679e-10 335-368 DM00215 19.43 3.136e-09 342-375 DM00215 19.43 5.119e-09 315-348 DM00215 19.43 8.322e-09 326-359
1728	PR00049	WILMS TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 8.627e-09 335-350
1728	BL00048	Protamine P1 proteins.	BL00048 6.39 5.026e-10 152-179 BL00048 6.39 6.329e-10 173-200 BL00048 6.39 8.224e-10 161-188 BL00048 6.39 3.363e-09 155-182 BL00048 6.39 3.475e-09 163-190 BL00048 6.39 3.925e-09 167-194 BL00048 6.39 4.150e-09 151-178 BL00048 6.39 4.150e-09 159-186 BL00048 6.39 4.825e-09 171-198 BL00048 6.39 5.838e-09 176-203 BL00048 6.39 8.200e-09 177-204 BL00048 6.39 9.550e-09 153-180
1728	PR00211	GLUTELIN SIGNATURE	PR00211B 0.86 1.827e-11 345-366 PR00211B 0.86 3.571e-11 339-360 PR00211B 0.86 6.917e-09 325-346 PR00211B 0.86 1.000e-08 351-372
1731	BL01221	PMP-22 / EMP / MP20 family proteins.	BL01221C 26.20 1.281e-34 59-104 BL01221D 13.99 5.966e-27 136-163 BL01221A 17.26 2.385e-26 1-29 BL01221B 13.29 1.000e-14 38-52
1733	BL00027	'Homeobox' domain proteins.	BL00027 26.43 4.000e-10 297-340
1733	PR00024	HOMEBOX SIGNATURE	PR00024A 11.87 4.150e-09 289-301
1734	BL00027	'Homeobox' domain proteins.	BL00027 26.43 4.000e-10 297-340
1734	PR00024	HOMEBOX SIGNATURE	PR00024A 11.87 4.150e-09 289-301
1738	BL00303	S-100/ICaBP type calcium binding protein.	BL00303B 26.15 5.075e-13 73-110
1738	BL00018	EF-hand calcium-binding domain proteins.	BL00018 7.41 3.400e-12 85-98 BL00018 7.41 8.043e-09 49-62
1738	PR00450	RECOVERIN FAMILY SIGNATURE	PR00450C 12.22 6.582e-09 44-66 PR00450C 12.22 9.772e-09 80-102
1740	PD01941	TRANSMEMBRANE COTRANSPORTER SYMP.	PD01941C 19.96 4.960e-16 84-139 PD01941B 15.02 2.093e-11 4-51
1742	BL00672	Serine proteases, V8 family, histidine proteins.	BL00672B 9.84 3.554e-09 214-231
1742	PR00839	V8 SERINE PROTEASE FAMILY SIGNATURE	PR00839E 12.04 8.062e-09 213-230
1745	BL00674	AAA-protein family proteins.	BL00674B 4.46 7.814e-10 360-382
1745	DM01022	LRR REPEAT.	DM01022A 7.35 1.900e-09 954-

SEQ ID NO:	Database entry ID	Description	Results*
			961
1745	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 6.186e-09 267-282
1746	BL00674	AAA-protein family proteins.	BL00674B 4.46 7.814e-10 360-382
1746	DM01022	LRR REPEAT.	DM01022A 7.35 1.900e-09 954-961
1746	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 6.186e-09 267-282
1747	BL00215	Mitochondrial energy transfer proteins.	BL00215A 15.82 9.591e-18 206-231 BL00215A 15.82 4.000e-15 104-129 BL00215A 15.82 9.400e-15 7-32 BL00215B 10.44 1.000e-10 154-167
1747	PR00927	ADENINE NUCLEOTIDE TRANSLOCATOR I SIGNATURE	PR00927G 11.07 9.036e-11 158-174 PR00927B 14.66 4.652e-10 239-261
1747	PR00926	MITOCHONDRIAL CARRIER PROTEIN SIGNATURE	PR00926F 17.75 2.826e-09 9-32 PR00926F 17.75 3.217e-09 208-231
1749	BL01013	Oxysterol-binding protein family proteins.	BL01013A 25.14 5.500e-21 537-573 BL01013D 26.81 2.161e-18 807-851 BL01013C 9.97 4.231e-13 625-635 BL01013B 11.33 3.017e-11 603-614
1751	BL00711	Lipoxygenases iron-binding region proteins.	BL00711I 18.56 8.630e-28 577-615 BL00711E 19.66 3.550e-22 414-451 BL00711G 21.83 9.100e-22 503-535 BL00711C 20.75 5.959e-19 268-297 BL00711D 17.56 1.923e-16 347-373 BL00711H 23.34 1.771e-12 535-574 BL00711F 19.79 2.086e-10 484-501
1751	PR00087	LIPOXYGENASE SIGNATURE	PR00087C 15.00 1.184e-17 423-444 PR00087A 18.37 7.061e-12 385-403 PR00087B 15.25 5.091e-10 403-421
1751	PR00467	MAMMALIAN LIPOXYGENASE SIGNATURE	PR00467E 9.00 3.400e-14 344-364 PR00467D 16.69 4.082e-09 243-265
1753	PR00492	RHO PROTEIN GDP DISSOCIATION INHIBITOR SIGNATURE	PR00492C 9.68 1.900e-23 122-139 PR00492B 9.77 8.579e-23 76-95 PR00492D 14.82 8.200e-21 139-155 PR00492A 11.92 1.643e-18 60-76
1756	BL00378	Hexokinases proteins.	BL00378A 19.01 8.500e-09 403-431
1757	BL00027	'Homeobox' domain proteins.	BL00027 26.43 8.615e-33 35-78
1757	BL00032	'Homeobox' antennapedia-type protein.	BL00032B 10.83 4.259e-27 24-63 BL00032C 11.28 5.909e-20 63-81
1757	PR00025	HOMEOTIC ANTENNAPEDIA PROTEIN SIGNATURE	PR00025B 11.94 4.000e-11 21-37
1757	PR00031	LAMBDA AND OTHER REPRESSOR HELIX-TURN-HELIX SIGNATURE	PR00031B 16.29 4.960e-11 58-75
1757	PR00024	HOMEBOX SIGNATURE	PR00024C 7.49 9.357e-13 67-77 PR00024B 11.27 3.500e-11 57-68 PR00024A 11.87 9.400e-11 42-54

SEQ ID NO:	Database entry ID	Description	Results*
1758	PR00179	LIPOCALIN SIGNATURE	PR00179B 9.56 1.000e-12 102-115 PR00179C 19.02 1.000e-10 130-146 PR00179A 13.78 5.680e-10 37-50
1758	BL00213	Lipocalin proteins.	BL00213B 8.78 8.000e-10 102-113 BL00213A 12.95 9.526e-10 37-51
1759	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 1.818e-11 164-187
1762	BL00269	Mammalian defensins proteins.	BL00269C 16.52 7.158e-09 171-200
1762	PD02283	PROTEIN SPORULATION REPEAT PRECU.	PD02283C 17.54 5.855e-10 57-85 PD02283C 17.54 5.855e-10 87-115 PD02283C 17.54 6.566e-10 117-145 PD02283C 17.54 1.450e-09 47-75 PD02283C 17.54 1.450e-09 77-105 PD02283C 17.54 1.450e-09 107-135 PD02283C 17.54 5.613e-09 67-95 PD02283C 17.54 5.613e-09 97-125 PD02283C 17.54 6.175e-09 137-165 PD02283C 17.54 7.525e-09 37-65 PD02283C 17.54 8.875e-09 147-175
1762	BL00203	Vertebrate metallothioneins proteins.	BL00203 13.94 3.379e-12 95-141 BL00203 13.94 3.690e-12 65-111 BL00203 13.94 2.978e-11 35-81 BL00203 13.94 5.549e-11 39-85 BL00203 13.94 6.538e-11 55-101 BL00203 13.94 6.538e-11 85-131 BL00203 13.94 7.231e-11 34-80 BL00203 13.94 7.429e-11 125-171 BL00203 13.94 7.527e-11 69-115 BL00203 13.94 8.220e-11 99-145 BL00203 13.94 1.670e-10 64-110 BL00203 13.94 2.053e-10 94-140 BL00203 13.94 2.149e-10 124-170 BL00203 13.94 2.819e-10 159-205 BL00203 13.94 5.213e-10 54-100 BL00203 13.94 5.213e-10 84-130 BL00203 13.94 5.691e-10 59-105 BL00203 13.94 5.691e-10 89-135 BL00203 13.94 6.936e-10 129-175 BL00203 13.94 7.511e-10 115-161 BL00203 13.94 7.702e-10 49-95 BL00203 13.94 7.702e-10 79-125 BL00203 13.94 1.551e-09 44-90 BL00203 13.94 1.551e-09 74-120 BL00203 13.94 3.112e-09 29-75 BL00203 13.94 4.031e-09 30-76 BL00203 13.94 4.214e-09 90-136 BL00203 13.94 4.306e-09 60-106 BL00203 13.94 5.133e-09 145-191 BL00203 13.94 6.235e-09 119-165 BL00203 13.94 6.327e-09 50-96 BL00203 13.94 6.327e-09 80-126 BL00203 13.94

SEQ ID NO:	Database entry ID	Description	Results*
			6.694e-09 45-91 BL00203 13.94 6.694e-09 75-121 BL00203 13.94 8.898e-09 104-150
1763	BL00216	Sugar transport proteins.	BL00216B 27.64 5.846e-09 141-191
1766	BL00456	Sodium:solute symporter family proteins.	BL00456A 22.59 2.080e-30 83-138 BL00456C 24.55 3.721e-29 221-276 BL00456B 18.94 1.000e-22 159-189
1766	PR00175	SODIUM/ALANINE SYMPORTER SIGNATURE	PR00175B 10.80 9.878e-09 226-245
1767	BL00142	Neutral zinc metalloproteinases, zinc-binding region proteins.	BL00142 8.38 1.857e-09 494-505
1768	BL00509	Ras GTPase-activating proteins.	BL00509B 10.28 1.643e-12 610-621
1772	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 6.143e-13 252-266 PR00048A 10.52 7.429e-13 476-490 PR00048A 10.52 3.118e-12 336-350 PR00048A 10.52 3.118e-12 364-378 PR00048A 10.52 4.706e-12 504-518 PR00048A 10.52 8.412e-12 224-238 PR00048A 10.52 3.842e-11 392-406 PR00048A 10.52 6.211e-11 308-322 PR00048A 10.52 6.211e-11 448-462 PR00048B 6.02 7.231e-11 492-502 PR00048B 6.02 3.250e-10 240-250 PR00048A 10.52 6.870e-10 420-434 PR00048B 6.02 2.421e-09 380-390
1772	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 8.800e-14 327-340 PD00066 13.92 1.500e-13 411-424 PD00066 13.92 5.500e-13 383-396 PD00066 13.92 5.500e-13 439-452 PD00066 13.92 7.500e-13 495-508 PD00066 13.92 9.000e-13 467-480 PD00066 13.92 3.571e-12 355-368 PD00066 13.92 7.000e-12 271-284 PD00066 13.92 7.923e-10 299-312 PD00066 13.92 2.500e-09 243-256
1772	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 3.250e-13 395-412 BL00028 16.07 5.950e-13 367-384 BL00028 16.07 6.478e-12 451-468 BL00028 16.07 8.435e-12 339-356 BL00028 16.07 1.692e-11 255-272 BL00028 16.07 3.769e-11 227-244 BL00028 16.07 5.154e-11 507-524 BL00028 16.07 2.200e-10 479-496 BL00028 16.07 9.400e-10 199-216 BL00028 16.07 2.029e-09 423-440 BL00028 16.07 3.571e-09 311-328
1773	PR00122	VACUOLAR ATP SYNTHASE 16 KD	PR00122D 9.97 7.214e-11 103-

SEQ ID NO:	Database entry ID	Description	Results*
		SUBUNIT SIGNATURE	127 PR00122C 8.20 9.526e-10 76-103
1773	BL00605	ATP synthase c subunit proteins.	BL00605 27.67 4.977e-09 70-124
1774	BL00518	Zinc finger, C3HC4 type (RING finger), proteins.	BL00518 12.23 6.571e-10 30-39
1776	BL01160	Kinesin light chain repeat proteins.	BL01160B 19.54 1.610e-09 33-87 BL01160B 19.54 9.619e-09 65-119
1783	PD02411	PROTEIN TRANSCRIPTION REGULATION NUCLEAR.	PD02411 21.89 6.786e-15 3967-4001
1783	PF00856	SET domain proteins.	PF00856B 16.42 6.595e-19 3949-3971 PF00856A 26.14 4.125e-12 3896-3933
1783	PF00628	PHD-finger.	PF00628 15.84 3.455e-12 86-101 PF00628 15.84 7.750e-10 38-53 PF00628 15.84 5.645e-09 164-179
1783	BL00115	Eukaryotic RNA polymerase II heptapeptide repeat proteins.	BL00115Z 3.12 4.971e-09 2575-2624 BL00115Z 3.12 7.750e-09 2582-2631
1784	DM01970	0 kw ZK632.12 YDR313C ENDOSOMAL III.	DM01970B 8.60 9.423e-10 111-124
1785	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 2.174e-10 177-191
1790	PF00922	Vesiculovirus phosphoprotein.	PF00922A 19.17 7.724e-09 276-310
1792	PD02059	CORE POLYPROTEIN PROTEIN GAG CONTAINS: P.	PD02059A 28.10 5.950e-10 34-75
1794	BL00326	Tropomyosins proteins.	BL00326D 8.76 8.065e-09 165-206
1795	BL00326	Tropomyosins proteins.	BL00326D 8.76 8.065e-09 173-214
1797	PR00563	BETA-3 ADRENERGIC RECEPTOR SIGNATURE	PR00563B 3.98 8.141e-09 8-28
1799	PR00450	RECOVERIN FAMILY SIGNATURE	PR00450C 12.22 1.570e-09 285-307
1801	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 8.800e-14 290-303 PD00066 13.92 4.000e-13 234-247 PD00066 13.92 4.429e-12 262-275 PD00066 13.92 9.217e-11 206-219 PD00066 13.92 3.769e-10 505-518 PD00066 13.92 4.115e-10 449-462 PD00066 13.92 4.462e-10 533-546 PD00066 13.92 6.538e-10 477-490
1801	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 5.091e-15 243-257 PR00048A 10.52 7.750e-14 542-556 PR00048A 10.52 3.647e-12 215-229 PR00048A 10.52 4.176e-12 486-500 PR00048B 6.02 6.000e-12 231-241 PR00048B 6.02 6.000e-12 287-297 PR00048A 10.52 7.353e-12 187-201 PR00048A 10.52 6.684e-11 271-285 PR00048A 10.52 4.130e-10 299-313 PR00048A 10.52 3.520e-09 430-444 PR00048A 10.52 3.880e-09 514-

SEQ ID NO:	Database entry ID	Description	Results*
			528 PR00048A 10.52 6.400e-09 458-472 PR00048A 10.52 8.560e-09 159-173
1801	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 1.563e-15 545-562 BL00028 16.07 1.450e-13 218-235 BL00028 16.07 3.348e-12 489-506 BL00028 16.07 7.652e-12 274-291 BL00028 16.07 2.385e-11 433-450 BL00028 16.07 4.115e-11 517-534 BL00028 16.07 5.154e-11 246-263 BL00028 16.07 1.000e-10 302-319 BL00028 16.07 5.200e-10 461-478 BL00028 16.07 6.700e-10 190-207 BL00028 16.07 1.257e-09 357-374 BL00028 16.07 9.486e-09 162-179
1802	BL00615	C-type lectin domain proteins.	BL00615A 16.68 8.920e-11 137-155
1802	PR00770	EOSINOPHIL MAJOR BASIC PROTEIN SIGNATURE	PR00770F 13.79 1.774e-09 198-218
1802	PR00356	TYPE II ANTIFREEZE PROTEIN SIGNATURE	PR00356B 14.85 2.521e-09 137-155
1803	BL00615	C-type lectin domain proteins.	BL00615A 16.68 8.920e-11 176-194
1803	PR00770	EOSINOPHIL MAJOR BASIC PROTEIN SIGNATURE	PR00770F 13.79 1.774e-09 237-257
1803	PR00356	TYPE II ANTIFREEZE PROTEIN SIGNATURE	PR00356B 14.85 2.521e-09 176-194
1804	BL00018	EF-hand calcium-binding domain proteins.	BL00018 7.41 9.000e-14 65-78
1806	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 1.000e-09 219-233
1808	PD02474	SYNTHASE SMALL SUBUNIT ACETOLACT.	PD02474B 21.08 8.568e-09 199-238
1809	PF00922	Vesiculovirus phosphoprotein.	PF00922A 19.17 1.000e-08 249-283
1812	PR00289	DISINTEGRIN SIGNATURE	PR00289B 11.79 1.947e-09 522-535
1814	PF00242	DNA polymerase (viral) N-terminal domain proteins.	PF00242F 12.18 8.522e-09 197-219
1815	PR00780	LEUSERPIN 2 SIGNATURE	PR00780B 4.89 4.491e-09 262-285
1816	BL00226	Intermediate filaments proteins.	BL00226D 19.10 8.027e-13 208-255
1817	PD01876	ANTIGEN MELANOMA-ASSOCIATED MULTIGENE FAMILY TUM.	PD01876C 21.73 3.326e-15 481-534 PD01876C 21.73 3.045e-10 735-788
1818	PR00747	GLYCOSYL HYDROLASE FAMILY 47 SIGNATURE	PR00747C 12.06 8.767e-09 337-356
1820	DM01782	HYDROGENASE (FE) LARGE CHAIN.	DM01782C 13.88 4.400e-19 349-368 DM01782F 9.01 4.375e-18 499-515 DM01782B 17.29 3.412e-10 294-327
1821	BL00226	Intermediate filaments proteins.	BL00226D 19.10 7.375e-38 321-368 BL00226B 23.86 7.107e-32 155-203 BL00226C 13.23 3.100e-

SEQ ID NO:	Database entry ID	Description	Results*
			19 220-251 BL00226A 12.77 7.000e-15 55-70 BL00226D 19.10 7.800e-09 254-301
1822	DM01415	6 SALIVARY GLUE PROTEIN.	DM01415B 13.78 9.518e-10 4-52
1822	BL00026	Chitin recognition or binding domain proteins.	BL00026 12.95 3.013e-09 43-64
1822	PD02283	PROTEIN SPORULATION REPEAT PRECU.	PD02283C 17.54 3.588e-09 110-138 PD02283C 17.54 3.588e-09 120-148
1822	PD00866	GLYCOPROTEIN PROTEIN SPIKE E2 PRECURSOR PEPLIMER.	PD00866L 3.73 1.443e-09 21-31 PD00866L 3.73 2.770e-09 97-107 PD00866L 3.73 2.770e-09 146-156 PD00866L 3.73 2.918e-09 7-17 PD00866L 3.73 2.918e-09 14-24 PD00866L 3.73 4.541e-09 4-14
1822	PR00858	CRUSTACEAN METALLOTHIONEIN SIGNATURE	PR00858B 5.93 2.819e-09 84-103 PR00858B 5.93 5.021e-09 114-133 PR00858B 5.93 5.021e-09 124-143
1822	PR00875	MOLLUSC METALLOTHIONEIN SIGNATURE	PR00875B 5.24 6.595e-09 85-93
1822	PR00874	FUNGI-IV METALLOTHIONEIN SIGNATURE	PR00874C 4.37 9.739e-10 73-88 PR00874C 4.37 7.250e-09 83-98
1822	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243I 31.77 3.143e-11 13-56 BL00243I 31.77 3.647e-10 26-69 BL00243I 31.77 6.426e-10 106-149 BL00243I 31.77 7.088e-10 96-139 BL00243I 31.77 9.338e-10 36-79 BL00243I 31.77 1.254e-09 46-89 BL00243I 31.77 6.451e-09 3-46 BL00243I 31.77 6.704e-09 77-120 BL00243I 31.77 7.211e-09 67-110 BL00243I 31.77 7.592e-09 116-159 BL00243I 31.77 8.606e-09 92-135
1822	BL00198	4Fe-4S ferredoxins, iron-sulfur binding region proteins.	BL00198 10.43 9.700e-09 6-18
1822	BL00203	Vertebrate metallothioneins proteins.	BL00203 13.94 8.024e-14 102-148 BL00203 13.94 7.750e-13 127-173 BL00203 13.94 1.103e-12 97-143 BL00203 13.94 3.172e-12 103-149 BL00203 13.94 3.379e-12 92-138 BL00203 13.94 4.207e-12 98-144 BL00203 13.94 4.207e-12 116-162 BL00203 13.94 5.345e-12 71-117 BL00203 13.94 5.345e-12 107-153 BL00203 13.94 9.897e-12 106-152 BL00203 13.94 1.791e-11 118-164 BL00203 13.94 2.879e-11 126-172 BL00203 13.94 3.176e-11 87-133 BL00203 13.94 4.758e-11 61-107 BL00203 13.94 5.846e-11 113-159 BL00203 13.94 6.044e-11 112-158 BL00203 13.94 7.231e-11 93-139 BL00203 13.94 1.287e-10

SEQ ID NO:	Database entry ID	Description	Results*
			128-174 BL00203 13.94 2.245e-10 76-122 BL00203 13.94 3.202e-10 23-69 BL00203 13.94 4.926e-10 78-124 BL00203 13.94 6.362e-10 81-127 BL00203 13.94 6.553e-10 82-128 BL00203 13.94 6.840e-10 111-157 BL00203 13.94 8.851e-10 43-89 BL00203 13.94 8.851e-10 96-142 BL00203 13.94 2.837e-09 8-54 BL00203 13.94 3.296e-09 72-118 BL00203 13.94 3.847e-09 117-163 BL00203 13.94 3.939e-09 86-132 BL00203 13.94 5.592e-09 123-169 BL00203 13.94 5.776e-09 108-154 BL00203 13.94 6.143e-09 30-76 BL00203 13.94 6.143e-09 79-125 BL00203 13.94 6.969e-09 121-167 BL00203 13.94 7.612e-09 16-62 BL00203 13.94 7.796e-09 101-147 BL00203 13.94 8.163e-09 33-79 BL00203 13.94 9.633e-09 77-123 BL00203 13.94 1.000e-08 66-112
1824	PR00860	VERTEBRATE METALLOTHIONEIN SIGNATURE	PR00860B 7.04 2.929e-20 74-88 PR00860A 5.46 5.655e-13 52-65 PR00860C 9.61 2.400e-12 88-98
1824	PR00858	CRUSTACEAN METALLOTHIONEIN SIGNATURE	PR00858B 5.93 1.419e-11 70-89 PR00858B 5.93 7.070e-11 65-84
1824	PR00874	FUNGI-IV METALLOTHIONEIN SIGNATURE	PR00874C 4.37 3.478e-10 64-79
1824	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 7.875e-10 59-85 BL00243I 31.77 4.803e-09 65-108
1824	PR00876	NEMATODE METALLOTHIONEIN SIGNATURE	PR00876D 5.77 2.191e-10 62-75 PR00876A 6.60 5.886e-09 61-74
1824	BL00203	Vertebrate metallothioneins proteins.	BL00203 13.94 1.000e-40 62-108 BL00203 13.94 7.429e-09 54-100 BL00203 13.94 8.071e-09 52-98 BL00203 13.94 8.806e-09 55-101
1824	PR00875	MOLLUSC METALLOTHIONEIN SIGNATURE	PR00875D 5.00 9.471e-09 59-70
1825	PR00360	C2 DOMAIN SIGNATURE	PR00360B 13.61 7.136e-09 572-586
1825	PR00399	SYNAPTOTAGMIN SIGNATURE	PR00399A 9.52 8.875e-09 360-376
1829	DM01206	CORONAVIRUS NUCLEOCAPSID PROTEIN.	DM01206B 10.69 8.767e-10 567-587 DM01206B 10.69 1.000e-09 563-583
1829	PD01351	PROTEIN REPEAT NEUROFILAMENT TRIPL.	PD01351B 13.72 6.786e-10 196-222 PD01351B 13.72 2.597e-09 198-224
1829	BL00035	'POU' domain proteins.	BL00035B 14.46 3.127e-09 634-655
1829	BL00229	Tau and MAP proteins tubulin-binding domain proteins.	BL00229A 23.57 3.182e-09 178-217
1829	PR00049	WILMS TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 3.647e-11 204-219 PR00049D 0.00 9.471e-11 209-224 PR00049D 0.00 8.500e-

SEQ ID NO:	Database entry ID	Description	Results*
			10 206-221 PR00049D 0.00 8.500e-10 207-222 PR00049D 0.00 3.746e-09 182-197
1829	BL00319	Amyloidogenic glycoprotein extracellular domain proteins.	BL00319C 17.12 3.132e-09 265-299 BL00319C 17.12 4.553e-09 1013-1047 BL00319C 17.12 5.618e-09 1019-1053 BL00319C 17.12 7.395e-09 267-301 BL00319C 17.12 7.632e-09 1017-1051
1829	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 3.089e-10 185-218 DM00215 19.43 8.393e-10 190-223 DM00215 19.43 2.373e-09 186-219 DM00215 19.43 7.102e-09 183-216 DM00215 19.43 8.169e-09 188-221
1829	BL00422	Granins proteins.	BL00422C 16.18 8.588e-09 263-291
1829	BL00412	Neuromodulin (GAP-43) proteins.	BL00412D 16.54 9.182e-11 1005-1056 BL00412D 16.54 9.120e-10 1004-1055 BL00412D 16.54 4.857e-09 1003-1054 BL00412D 16.54 8.347e-09 1008-1059 BL00412D 16.54 9.449e-09 1001-1052
1829	PR00832	PAXILLIN SIGNATURE	PR00832B 9.87 9.526e-09 377-401
1829	PD02059	CORE POLYPROTEIN PROTEIN GAG CONTAINS: P.	PD02059B 24.48 9.620e-09 196-231
1829	PF00992	Troponin.	PF00992A 16.67 9.882e-09 1005-1040
1829	BL00048	Protamine P1 proteins.	BL00048 6.39 6.949e-15 569-596 BL00048 6.39 1.885e-14 568-595 BL00048 6.39 3.361e-14 570-597 BL00048 6.39 8.377e-14 577-604 BL00048 6.39 8.377e-14 578-605 BL00048 6.39 3.631e-13 571-598 BL00048 6.39 4.738e-13 576-603 BL00048 6.39 7.369e-13 582-609 BL00048 6.39 2.456e-12 575-602 BL00048 6.39 3.118e-12 573-600 BL00048 6.39 3.515e-12 567-594 BL00048 6.39 5.235e-12 581-608 BL00048 6.39 5.632e-12 583-610 BL00048 6.39 7.221e-12 557-584 BL00048 6.39 2.250e-11 572-599 BL00048 6.39 2.875e-11 585-612 BL00048 6.39 4.875e-11 586-613 BL00048 6.39 5.375e-11 591-618 BL00048 6.39 7.375e-11 589-616 BL00048 6.39 7.500e-11 580-607 BL00048 6.39 8.625e-11 588-615 BL00048 6.39 2.895e-10 563-590 BL00048 6.39 5.382e-10 574-601 BL00048 6.39 5.500e-10 566-593 BL00048 6.39 5.618e-10 584-611 BL00048 6.39 9.171e-10 561-588 BL00048 6.39 9.882e-10 592-619

SEQ ID NO:	Database entry ID	Description	Results*
			BL00048 6.39 1.450e-09 590-617 BL00048 6.39 3.925e-09 560-587 BL00048 6.39 4.488e-09 562-589 BL00048 6.39 4.938e-09 579-606 BL00048 6.39 5.275e-09 564-591 BL00048 6.39 5.725e-09 558-585 BL00048 6.39 5.725e-09 587-614 BL00048 6.39 6.625e-09 555-582 BL00048 6.39 7.075e-09 556-583 BL00048 6.39 9.438e-09 559-586 BL00048 6.39 9.888e-09 600-627
1829	PD01234	PROTEIN NUCLEAR BROMODOMAIN TRANS.	PD01234B 15.53 1.000e-08 201-219
1830	BL00092	N-6 Adenine-specific DNA methylases proteins.	BL00092 5.35 2.000e-09 136-145
1831	PR00511	TEKTN SIGNATURE	PR00511A 13.59 3.700e-14 113-130
1833	PR00764	COMPLEMENT C9 SIGNATURE	PR00764F 16.89 2.286e-09 158-179
1833	BL00022	EGF-like domain proteins.	BL00022B 7.54 4.600e-09 138-145
1833	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 4.086e-10 167-183 BL01187B 12.04 4.600e-09 104-120
1833	BL01185	C-terminal cystine knot proteins.	BL01185B 21.14 6.929e-09 146-195
1833	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010C 11.16 7.429e-09 109-120 PR00010C 11.16 7.643e-09 172-183 PR00010A 11.79 8.846e-09 190-202
1833	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011A 14.06 4.822e-09 208-227 PR00011D 14.03 8.957e-09 67-86
1833	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243I 31.77 1.000e-08 34-77
1835	BL01279	Protein-L-isopartate(D-aspartate) O-methyltransferase signa.	BL01279A 24.27 3.691e-09 419-467
1835	BL01131	Ribosomal RNA adenine dimethylases proteins.	BL01131A 26.62 4.600e-09 421-467
1835	BL01230	RNA methyltransferase trmA family proteins.	BL01230E 15.79 6.607e-11 571-587 BL01230A 17.88 8.962e-10 409-428 BL01230B 11.62 8.475e-09 436-449
1835	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 8.780e-09 598-613
1837	PD01719	PRECURSOR GLYCOPROTEIN SIGNAL RE.	PD01719A 12.89 2.603e-11 259-287 PD01719A 12.89 8.105e-10 199-227
1838	BL01162	Quinone oxidoreductase / zeta-crystallin proteins.	BL01162C 22.80 1.269e-18 151-195 BL01162A 15.38 1.265e-11 64-87
1838	BL00279	Membrane attack complex components / perforin proteins.	BL00279C 31.64 3.156e-09 134-188
1838	BL00059	Zinc-containing alcohol dehydrogenases proteins.	BL00059B 16.08 7.273e-09 93-121
1841	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 5.415e-26 46-85

SEQ ID NO:	Database entry ID	Description	Results*
1841	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 5.737e-11 140-154 PR00048A 10.52 6.087e-10 224-238
1841	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 4.130e-12 227-244 BL00028 16.07 7.577e-11 352-369 BL00028 16.07 3.400e-10 380-397 BL00028 16.07 6.400e-10 199-216 BL00028 16.07 1.257e-09 143-160 BL00028 16.07 2.029e-09 171-188 BL00028 16.07 5.886e-09 408-425
1841	PR00967	ACUTE MYELOID LEUKEMIA 1 PROTEIN SIGNATURE	PR00967I 12.41 8.130e-09 466-481
1841	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 2.038e-10 187-200 PD00066 13.92 1.600e-09 396-409 PD00066 13.92 9.400e-09 215-228
1841	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479A 19.86 6.553e-09 214-237 BL00479A 19.86 9.809e-09 367-390
1842	PD02910	TRANSCRIPTION PROTEIN FACTOR REGULATION A.	PD02910A 15.43 9.839e-09 62-97
1843	PD02199	SUBUNIT HYDROGEN ION TRANSPORT T.	PD02199A 20.58 1.000e-40 10-61 PD02199D 13.18 1.000e-40 364-405 PD02199F 15.02 1.000e-40 440-482 PD02199J 11.42 1.000e-40 723-762 PD02199K 15.22 1.000e-40 792-831 PD02199G 9.43 4.447e-24 531-555 PD02199B 27.90 1.474e-22 263-306 PD02199H 13.62 2.636e-21 576-599 PD02199E 7.56 8.642e-19 405-424 PD02199C 17.60 8.085e-14 313-329 PD02199I 8.90 4.780e-09 616-624
1844	BL00218	Amino acid permeases proteins.	BL00218E 23.30 5.920e-10 343-383
1845	BL00048	Protamine P1 proteins.	BL00048 6.39 9.526e-10 160-187
1845	BL00422	Granins proteins.	BL00422C 16.18 4.000e-09 590-618
1845	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 8.385e-09 943-958
1845	PF00992	Troponin.	PF00992A 16.67 7.900e-13 568-603 PF00992A 16.67 4.090e-11 566-601 PF00992A 16.67 5.817e-10 570-605 PF00992A 16.67 8.479e-10 579-614 PF00992A 16.67 2.066e-09 564-599 PF00992A 16.67 4.789e-09 575-610 PF00992A 16.67 4.908e-09 532-567 PF00992A 16.67 6.803e-09 536-571 PF00992A 16.67 7.632e-09 562-597 PF00992A 16.67 8.697e-09 585-620 PF00992A 16.67 9.053e-09 583-618 PF00992A 16.67 9.289e-09 516-551

SEQ ID NO:	Database entry ID	Description	Results*
1845	PF01140	Matrix protein (MA), p15.	PF01140D 15.54 8.500e-12 505-540 PF01140D 15.54 7.120e-11 584-619 PF01140D 15.54 9.760e-11 586-621 PF01140D 15.54 3.813e-10 588-623 PF01140D 15.54 4.938e-10 563-598 PF01140D 15.54 6.738e-10 519-554 PF01140D 15.54 8.313e-10 503-538 PF01140D 15.54 9.325e-10 549-584 PF01140D 15.54 9.663e-10 567-602 PF01140D 15.54 9.775e-10 565-600 PF01140D 15.54 1.000e-09 582-617 PF01140D 15.54 2.884e-09 575-610 PF01140D 15.54 3.198e-09 551-586 PF01140D 15.54 3.302e-09 517-552 PF01140D 15.54 4.140e-09 533-568 PF01140D 15.54 4.872e-09 578-613 PF01140D 15.54 6.860e-09 581-616 PF01140D 15.54 7.174e-09 594-629 PF01140D 15.54 7.384e-09 576-611 PF01140D 15.54 7.593e-09 572-607 PF01140D 15.54 8.640e-09 570-605 PF01140D 15.54 8.744e-09 507-542 PF01140D 15.54 8.744e-09 596-631 PF01140D 15.54 9.163e-09 577-612 PF01140D 15.54 9.267e-09 579-614 PF01140D 15.54 9.791e-09 574-609 PF01140D 15.54 1.000e-08 531-566
1848	BL00811	Oleosins proteins.	BL00811B 10.57 9.791e-09 307-336
1852	BL00415	Synapsins proteins.	BL00415N 4.29 4.153e-09 301-345
1852	DM00668	ZEIN.	DM00668B 22.01 8.018e-09 291-343
1853	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 3.250e-17 133-164
1853	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109E 14.41 3.045e-11 254-277
1853	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 2.317e-16 689-744 PF00791B 28.49 3.753e-15 523-578 PF00791B 28.49 4.316e-12 656-711 PF00791B 28.49 1.727e-11 589-644 PF00791B 28.49 3.636e-11 556-611 PF00791C 20.98 4.913e-11 570-609 PF00791B 28.49 6.330e-10 722-777 PF00791C 20.98 5.853e-09 703-742
1853	PF00023	Ank repeat proteins.	PF00023A 16.03 5.200e-13 722-738 PF00023B 14.20 1.000e-12 652-662 PF00023A 16.03 2.000e-12 755-771 PF00023A 16.03 7.857e-11 656-672 PF00023A

SEQ ID NO:	Database entry ID	Description	Results*
			16.03 8.286e-11 622-638 PF00023B 14.20 4.682e-09 519-529 PF00023A 16.03 6.143e-09 589-605 PF00023A 16.03 6.786e-09 689-705
1853	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 9.000e-11 582-595 PD00078B 13.14 8.435e-09 649-662
1854	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 4.971e-14 4-26
1856	BL00317	WAP-type 'four-disulfide core' domain proteins.	BL00317B 14.58 3.550e-13 48-70
1856	PR00003	4-DISULPHIDE CORE SIGNATURE	PR00003C 7.69 6.357e-09 54-64
1857	BL01019	ADP-ribosylation factors family proteins.	BL01019B 19.49 7.517e-21 95-150
1857	BL01020	SAR1 family proteins.	BL01020C 15.35 2.301e-18 79-130
1857	PR00328	GTP-BINDING SAR1 PROTEIN SIGNATURE	PR00328C 13.16 2.841e-10 78-104
1858	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 2.385e-15 128-141 PD00066 13.92 5.714e-12 100-113
1858	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048B 6.02 7.188e-10 97-107 PR00048B 6.02 3.842e-09 125-135 PR00048A 10.52 6.040e-09 137-151
1859	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 1.000e-11 174-188 PR00048B 6.02 1.692e-11 162-172
1859	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 3.739e-11 165-178
1859	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 2.385e-11 177-194 BL00028 16.07 3.769e-11 121-138 BL00028 16.07 8.269e-11 149-166
1860	PF00651	BTB (also known as BR-C/Tik) domain proteins.	PF00651 15.00 2.895e-11 45-58
1860	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 2.038e-10 419-432 PD00066 13.92 7.231e-10 591-404
1860	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 5.696e-10 400-414 PR00048A 10.52 8.435e-10 428-442
1860	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 8.269e-11 403-420 BL00028 16.07 3.400e-10 375-392 BL00028 16.07 3.057e-09 431-448
1861	DM00547	1 kw CHROMO BROMODOMAIN SHADOW GLOBAL.	DM00547F 23.43 7.643e-34 606-653 DM00547B 11.28 7.907e-16 155-169 DM00547C 17.30 8.650e-14 209-231 DM00547D 11.60 6.500e-13 277-291 DM00547E 13.94 1.000e-11 307-330
1861	BL00039	DEAD-box subfamily ATP-dependent helicases proteins.	BL00039D 21.67 6.379e-10 590-636
1862	BL01115	GTP-binding nuclear protein ran proteins.	BL01115A 10.22 5.330e-11 18-62

SEQ ID NO:	Database entry ID	Description	Results*
1862	BL01019	ADP-ribosylation factors family proteins.	BL01019A 13.20 1.809e-09 52-92
1862	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449C 17.27 3.647e-19 59-82 PR00449A 13.20 7.000e-15 18-40 PR00449D 10.79 8.875e-14 121-135 PR00449E 13.50 8.920e-14 157-180 PR00449B 14.34 8.500e-09 41-58
1867	BL01283	T-box domain proteins.	BL01283D 11.70 7.868e-31 59-92 BL01283C 13.05 2.537e-14 25-39
1867	PR00937	T-BOX DOMAIN SIGNATURE	PR00937D 13.41 5.378e-15 24-39 PR00937F 12.53 1.450e-12 83-92 PR00937E 11.86 5.592e-12 62-76 PR00937C 10.51 5.219e-10 5-15
1870	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803A 10.51 8.699e-09 100-121
1872	BL00470	Isocitrate and isopropylmalate dehydrogenases proteins.	BL00470A 16.25 5.179e-14 10-31 BL00470C 15.43 4.103e-10 223-238 BL00470E 16.52 1.900e-09 287-297
1873	PF00023	Ank repeat proteins.	PF00023A 16.03 3.893e-09 44-60 PF00023B 14.20 9.182e-09 40-50
1874	PF00023	Ank repeat proteins.	PF00023A 16.03 3.893e-09 72-88 PF00023B 14.20 9.182e-09 68-78
1877	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 2.019e-26 51-90
1877	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 6.786e-16 427-444 BL00028 16.07 1.900e-13 287-304 BL00028 16.07 3.700e-13 481-498 BL00028 16.07 1.000e-12 315-332 BL00028 16.07 1.000e-12 399-416 BL00028 16.07 3.348e-12 453-470 BL00028 16.07 4.522e-12 371-388 BL00028 16.07 6.885e-11 343-360 BL00028 16.07 4.600e-10 509-526
1877	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 8.500e-17 424-438 PR00048A 10.52 7.000e-14 312-326 PR00048A 10.52 1.643e-13 396-410 PR00048A 10.52 3.571e-13 478-492 PR00048B 6.02 9.000e-12 300-310 PR00048A 10.52 1.000e-11 506-520 PR00048A 10.52 5.737e-11 340-354 PR00048A 10.52 1.391e-10 284-298 PR00048B 6.02 6.063e-10 412-422 PR00048B 6.02 1.474e-09 494-504 PR00048B 6.02 2.895e-09 356-366
1877	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 4.600e-14 415-428 PD00066 13.92 7.000e-14 469-482 PD00066 13.92 5.500e-13 303-316 PD00066 13.92 4.429e-12 331-344 PD00066 13.92 9.217e-11 497-510 PD00066 13.92 2.038e-10 387-400 PD00066 13.92 6.400e-09

SEQ ID NO:	Database entry ID	Description	Results*
			359-372
1878	PR00830	ENDOPEPTIDASE LA (LON) SERINE PROTEASE (S16) SIGNATURE	PR00830A 8.41 4.927e-13 222-242
1878	PR00300	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT SIGNATURE	PR00300A 9.56 1.545e-11 218-237
1878	PR00051	BACTERIAL CHROMOSOMAL REPLICATION INITIATOR (DNAA) SIGNATURE	PR00051A 10.68 3.647e-10 215-236
1878	PR00819	CBXX/CFQX SUPERFAMILY SIGNATURE	PR00819B 10.83 9.213e-10 217-233
1878	BL00674	AAA-protein family proteins.	BL00674B 4.46 9.000e-24 215-237 BL00674C 22.60 8.448e-20 248-291 BL00674D 23.41 5.140e-18 308-355 BL00674E 15.24 9.217e-16 390-410 BL00674A 16.91 5.304e-09 181-202
1878	BL00113	Adenylate kinase proteins.	BL00113A 12.74 8.615e-09 219-236
1879	BL00456	Sodium:solute symporter family proteins.	BL00456A 22.59 1.957e-32 35-90 BL00456B 18.94 9.780e-17 111-141
1880	BL00456	Sodium:solute symporter family proteins.	BL00456A 22.59 1.957e-32 35-90 BL00456C 24.55 1.225e-31 173-228 BL00456B 18.94 9.780e-17 111-141
1884	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 1.818e-11 159-182
1885	PR00830	ENDOPEPTIDASE LA (LON) SERINE PROTEASE (S16) SIGNATURE	PR00830A 8.41 5.897e-10 352-372
1885	BL00847	MCM family proteins.	BL00847D 15.16 8.568e-25 343-384 BL00847B 24.76 8.971e-25 194-237 BL00847C 18.79 9.270e-10 301-335
1885	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 1.931e-09 346-368
1885	BL00674	AAA-protein family proteins.	BL00674B 4.46 5.378e-09 345-367
1886	PF00651	BTB (also known as BR-C/Ttk) domain proteins.	PF00651 15.00 7.000e-10 50-63
1886	PR00501	KELCH REPEAT SIGNATURE	PR00501A 8.25 5.295e-09 507-521
1886	DM00099	4 kw A53R REDUCTASE TERMINAL DIHYDROPTERIDINE.	DM00099B 14.73 6.625e-09 479-489
1887	PD00306	PROTEIN GLYCOPROTEIN PRECURSOR RE.	PD00306A 10.26 6.625e-13 544-558
1887	PD01364	MUCIN GLYCOPROTEIN PRECURSOR MEM.	PD01364B 13.94 5.500e-09 544-560
1888	PD00306	PROTEIN GLYCOPROTEIN PRECURSOR RE.	PD00306A 10.26 6.625e-13 544-558
1888	PD01364	MUCIN GLYCOPROTEIN PRECURSOR MEM.	PD01364B 13.94 5.500e-09 544-560
1889	PD00306	PROTEIN GLYCOPROTEIN PRECURSOR RE.	PD00306A 10.26 6.625e-13 544-558
1889	PD01364	MUCIN GLYCOPROTEIN PRECURSOR MEM.	PD01364B 13.94 5.500e-09 544-560
1890	PF00938	Lipoprotein.	PF00938E 19.50 6.096e-09 272-

SEQ ID NO:	Database entry ID	Description	Results*
			307
1891	PF00925	GTP cyclohydrolase II.	PF00925F 13.23 9.850e-09 356-367
1893	BL00226	Intermediate filaments proteins.	BL00226A 12.77 5.355e-13 139-154
1895	PF00035	Double-stranded RNA binding motif.	PF00035B 12.08 7.750e-09 273-287
1896	PF00622	Domain in SP1a and the RYanodine Receptor.	PF00622B 21.00 9.250e-11 170-192

TABLE 4

SEQ ID	Model	Description	E-value	Score	Repeats	Position
950	UCH-2	Ubiquitin carboxyl-terminal hydrolase family	6.6e-26	99.5	1	825-886
950	UCH-1	Ubiquitin carboxyl-terminal hydrolases famil	2.5e-15	64.4	1	266-297
952	ethand	EF hand	0.0034	24.0	1	322-350
952	Adeno_E1A	Early E1A protein	8.9	-168.3	1	298-448
953	SH2	SH2 domain	1.5e-16	68.4	1	320-396
954	SH2	SH2 domain	1.5e-16	68.4	1	347-423
955	RCC1	Regulator of chromosome condensation (RCC1)	1.3e-13	58.6	4	148-197;200-249;318-367;370-418
958	BRCT	BRCA1 C Terminus (BRCT) domain	7.2	1.2	1	45-130
962	UQ_con	Ubiquitin-conjugating enzyme	5.4e-39	143.0	1	2-117
963	UQ_con	Ubiquitin-conjugating enzyme	1.6e-57	204.5	1	2-132
971	kinesin	Kinesin motor domain	2.2e-154	526.3	1	47-372
971	WD40	WD domain, G-beta repeat	1e-53	191.9	7	1327-1361;1367-1402;1432-1466;1472-1511;1523-1557;1564-1600;1606-1640
971	filament	Intermediate filament protein	2.6	-195.6	1	423-805
971	HdV_ag	Hepatitis delta virus delta antigen	6.2	-47.5	1	703-880
971	PFEMP	Plasmodium falciparum erythrocyte membrane p	8.9	-86.8	1	475-585
971	G6PD	Glucose-6-phosphate dehydrogenase, NAD bindi	9.6	-123.8	1	912-1049
971	DUF232	Putative transcriptional regulator	9.7	-30.1	1	616-750
975	Kelch	Kelch motif	5.7e-62	219.3	4	267-312;314-359;361-406;408-453
975	BTB	BTB/POZ domain	3.1e-38	140.4	1	23-130
977	kinesin	Kinesin motor domain	9.7e-	520.8	1	53-353

SEQ ID	Model	Description	E-value	Score	Repeats	Position
			153			
977	FHA	FHA domain	0.078	16.9	1	470-534
977	KMP11	Kinetoplastid membrane protein 11	3.5	-3.2	1	355-438
977	SART-1	SART-1 family	7.1	-360.1	1	325-881
977	Transposase 22	L1 transposable element	8.7	-180.7	1	322-622
979	Ribosomal L10	Ribosomal protein L10	6	-33.3	1	70-173
980	Aa trans oxidored_ql	Transmembrane amino acid transporter NADH-Ubiquinone/plastoquinone	1.6e-75	264.3	1	69-479
980	xan_ur_permease	Permease family	8.4	-201.3	1	137-471
980	Trp_Tyr_perm	Tryptophan/tyrosine permease family	9.6	-297.9	1	70-474
982	HYR	HYR domain	2e-35	131.1	2	106-187;188-267
982	EGF	EGF-like domain	3.9e-24	93.6	3	742-773;780-811;818-849
982	sushi	Sushi domain (SCR repeat)	0.28	12.5	3	1-38;43-104;272-332
982	TNFR_c6	TNFR/NGFR cysteine-rich region	0.74	9.6	1	563-601
982	laminin_EGF	Laminin EGF-like (Domains III and V)	6.6	-12.8	1	746-791
982	metallothio	Metallothionein	7.6	-11.6	1	744-804
982	HMG_CoA_synt	Hydroxymethylglutaryl-coenzyme A synthase	8.7	-342.7	1	336-625
986	MHC_II_alpha	Class II histocompatibility antigen, alpha	8.8e-13	55.9	1	37-106
986	ig	Immunoglobulin domain	8e-05	29.5	1	122-186
987	LRR	Leucine Rich Repeat	2.6e-12	54.3	4	68-91;92-114;115-137;138-159
987	UVR	UvrB/uvrC motif	5.3	-1.5	1	453-486
990	UVR	UvrB/uvrC motif	3.8	-0.3	1	552-588
991	CK_II_beta	Casein kinase II regulatory subunit	4.5e-69	242.9	1	5-124
994	RNA_pol_B	RNA polymerase beta subunit	0	1199.4	1	26-1010
994	PHD	PHD-finger	5.9	-17.0	1	1013-1048
995	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	4.9e-05	30.2	1	10-59
995	zf-B_box	B-box zinc finger	0.0022	24.7	1	92-134
996	HLH	Helix-loop-helix DNA-binding domain	2e-07	38.1	1	276-327
997	ras	Ras family	3.3e-12	2.0	1	23-145
998	pkinase	Protein kinase domain	3.1e-08	-16.4	1	1-139
1000	ig	Immunoglobulin domain	2.7e-06	34.4	2	42-95;225-281
1001	Y_phosphatase	Protein-tyrosine phosphatase	4.2	-86.2	1	180-409
1002	PX	PX domain	5.8e-19	76.4	1	22-138
1002	Peptidase_S21	Assemblin (Peptidase family S21)	8.9	-173.6	1	76-352
1003	Y_phosph	Protein-tyrosine phosphatase	1.1	-79.3	1	98-327

SEQ ID	Model	Description	E-value	Score	Repeats	Position
	atase					
1004	Y_phosphatase	Protein-tyrosine phosphatase	4.2	-86.2	1	180-409
1008	oxidored_q1	NADH:Ubiquinone/plastoquinone (complex I)	5.2	-165.9	1	64-311
1010	filament	Intermediate filament protein	1.7	-190.5	1	331-647
1010	Tektin	Tektin family	1.8	-228.5	1	192-507
1010	bZIP	bZIP transcription factor	4.4	-3.7	1	253-317
1010	spectrin	Spectrin repeat	5.6	-18.0	1	320-429
1010	SART-1	SART-1 family	8	-362.0	1	54-675
1010	Myosin_tail	Myosin tail	9.8	-555.1	1	6-734
1013	Defensin_propep	Defensin propeptide	1.2e-26	102.0	1	45-97
1013	defensins	Mammalian defensin	7.3e-14	59.5	1	110-138
1014	filament	Intermediate filament protein	0.69	-180.5	1	314-579
1014	PolyA_polymerase	Poly A polymerase family	2.5	-64.5	1	348-463
1014	ERM	Ezrin/radixin/moesin family	8.7	-223.3	1	226-487
1014	Transposase_12	Transposase	9.4	-152.0	1	155-465
1015	zf-C2H2	Zinc finger, C2H2 type	1.2e-55	198.2	13	129-152:349-371:379-401:407-429:446-468:474-496:505-527:533-556:562-585:903-925:931-953:959-981:987-1010
1015	60s_ribosomal	60s Acidic ribosomal protein	0.23	-21.0	1	61-194
1015	TFIIIS	Transcription factor S-II (TFIIIS)	0.82	2.1	1	446-484
1015	Rubredoxin	Rubredoxin	2.8	-8.3	1	900-943
1015	zf-BED	BED zinc finger	9	-7.0	1	972-1011
1021	SSF	Sodium:solute symporter family	1.7e-05	-65.8	1	5-184
1028	zf-C2H2	Zinc finger, C2H2 type	4.3e-30	113.4	5	100-122:132-154:160-182:188-210:216-238
1028	KRAB	KRAB box	7.1e-24	92.8	1	4-44
1028	zf-BED	BED zinc finger	0.63	3.4	1	78-123
1029	fn3	Fibronectin type III domain	1.8e-13	58.2	1	146-231
1029	Ig	Immunoglobulin domain	0.0013	25.5	1	275-335
1031	polyprenyl_synt	Polyprenyl synthetase	0.013	-82.6	1	119-332
1032	Defensin_propep	Defensin propeptide	1.2e-26	102.0	1	68-120
1032	defensins	Mammalian defensin	7.3e-14	59.5	1	133-161
1033	MAGE	MAGE family	3.8e-34	126.8	1	1-208

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1034	LRR	Leucine Rich Repeat	2.3e-17	71.1	6	62-84:87-110:111-134:135-158:159-181:186-209
1034	LRRCT	Leucine rich repeat C-terminal domain	7.4e-05	29.6	1	221-271
1034	ig	Immunoglobulin domain	0.061	19.9	1	283-343
1034	LRRNT	Leucine rich repeat N-terminal domain	0.19	18.3	1	33-60
1035	CoAe	Dephospho-CoA kinase	3e-93	323.2	1	359-537
1035	Cytidylyltransferase	Cytidylyltransferase	8.3e-06	31.4	1	191-315
1035	SKI	Shikimate kinase	0.64	-65.2	1	356-510
1035	ArgK	ArgK protein	7.5	-212.4	1	341-541
1038	lipase	Lipase	1.1e-12	49.8	1	1-198
1044	homeobox	Homeobox domain	2.6e-30	114.1	1	155-211
1048	SKI	Shikimate kinase	0.49	-63.6	1	6-185
1049	fn3	Fibronectin type III domain	4.7e-78	272.7	5	159-245:257-343:360-459:480-565:577-665
1049	ig	Immunoglobulin domain	2.5e-05	31.1	1	79-137
1050	ig	Immunoglobulin domain	0.019	21.5	2	1-58:93-142
1051	MHC_1	Class I Histocompatibility antigen, domains	2.1	-83.5	1	24-116
1053	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.88	0.8	1	198-232
1057	p450	Cytochrome P450	5.7e-08	-1.4	1	66-377
1060	Ribosomal S21	Ribosomal protein S21	0.5	-7.2	1	753-810
1060	Tropomyosin	Tropomyosin	9	-120.2	1	208-412
1064	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.88	0.8	1	141-175
1065	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.88	0.8	1	191-225
1066	aa_permeases	Amino acid permease	3.2e-06	-173.3	1	46-455
1066	Aa_trans	Transmembrane amino acid transporter	0.4	-172.1	1	48-434
1066	Trp_Tyr_perm	Tryptophan/tyrosine permease family	2.6	-282.5	1	48-392
1066	oxidored_q1	NADH-Ubiquinone/plastoquinone	4.6	-164.9	1	170-421
1066	xan_ur_pe rmense	Permease family	9.7	-202.9	1	60-397
1068	zf-C2H2	Zinc finger, C2H2 type	6.8e-125	428.3	22	101-128:169-191:197-219:225-247:253-275:281-303:309-324:330-352:358-380:386-408:414-436:442-464:470-

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						492:524- 546:552- 565:571- 593:599- 621:627- 649:655- 677:683- 705:711- 733:739- 761
1068	KRAB	KRAB box	4.1e-25	96.9	1	4-44
1068	GATA	GATA zinc finger	2.6	-7.8	1	279-325
1068	LIM	LIM domain	4.3	-17.1	1	311-368
1068	zf-BED	BED zinc finger	4.7	-4.4	1	315-353
1068	zf-TRAF	TRAF-type zinc finger	4.9	-6.5	1	163-212
1068	FYVE	FYVE zinc finger	6.7	-24.5	1	245-320
1068	TFIIS	Transcription factor S-II (TFIIS)	7.3	-6.0	1	281-319
1070	C2	C2 domain	1.1e-38	142.0	2	178- 263:319- 406
1071	ig	Immunoglobulin domain	1.9	12.9	1	19-66
1075	WD40	WD domain, G-beta repeat	1.7e-27	104.8	6	12-53:59- 96:160- 194:200- 236:245- 281:287- 322
1078	WD40	WD domain, G-beta repeat	2.2e-25	97.8	5	16-52:60- 96:106- 142:191- 227:234- 270
1078	FYVE	FYVE zinc finger	2e-22	87.9	1	279-353
1078	DnaI CX XCXGX G	DnaI central domain (4 repeats)	3.1	-45.6	1	304-360
1079	PH	PH domain	1.5e-25	98.3	2	120- 215:298- 391
1081	KRAB	KRAB box	6e-21	83.0	1	45-81
1082	Ribosoma l L11	Ribosomal protein L11	1e-64	228.4	1	117-248
1084	zf-C2H2	Zinc finger, C2H2 type	6.4e- 185	627.7	25	109- 131:137- 159:165- 187:221- 243:249- 271:277- 299:305- 327:333- 355:361- 383:389- 411:445- 467:473- 495:501- 523:529- 551:557- 579:585- 607:641-

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						663:669- 691:697- 719:753- 775:781- 803:809- 831:837- 859:865- 887:893- 915
1084	CBM_1	Fungal cellulose binding domain	2.6	4.4	1	38-66
1084	LIM	LIM domain	3	-15.7	1	783-847
1084	zF-BED	BED zinc finger	5.9	-5.3	2	205- 244:737- 776
1085	Band_7	SPFH domain / Band 7 family	8.1e-42	152.3	1	39-214
1087	TPR	TPR Domain	4.5e-16	66.8	2	58-91:92- 125
1090	WH1	WH1 domain	0.0017	11.6	1	11-119
1091	zF-DHHC	DHHC zinc finger domain	0.033	-11.5	1	120-158
1094	Calx-beta	Calx-beta domain	0.19	-11.6	1	23-117
1095	zF-C2H2	Zinc finger, C2H2 type	5.4e-82	285.8	12	288- 311:337- 359:365- 387:393- 415:421- 443:449- 471:477- 499:505- 527:533- 555:561- 583:589- 611:617- 639
1095	SCAN	SCAN domain	1.5e-54	194.6	1	46-141
1095	zF-BED	BED zinc finger	3.3	-3.0	2	434- 472:574- 612
1097	7tm_2	7 transmembrane receptor (Secretin family)	6.8e-21	82.8	1	325-580
1097	GPS	Latrophilin/CL-1-like GPS domain	9.5e-13	55.8	1	273-323
1097	Srg	C.elegans Srg family integral membrane prote	4.5	-217.5	1	309-365
1099	lectin_c	Lectin C-type domain	0.0011	7.2	1	6-100
1100	PDZ	PDZ domain (Also known as DHR or GLGF)	0.0014	25.3	1	12-91
1100	Tymo_45 kD_70kD	Tymovirus 45/70Kd protein	1.8	-283.9	1	1-398
1101	cadherin	Cadherin domain	8.9e-95	328.3	5	64- 156:170- 265:279- 381:394- 485:498- 595
1101	Cadherin_C term	Cadherin cytoplasmic region	4.7e-80	279.4	1	643-794
1103	COesterase	Carboxylesterase	0.98	-265.9	1	31-265
1104	DSPc	Dual specificity phosphatase, catalytic	9.7e-30	112.2	1	133-315

SEQ ID	Model	Description	E-value	Score	Repeats	Position
		doma				
1105	efhand	EF hand	3.1e-21	84.0	3	124-152:160-188:208-236
1108	zf-C2H2	Zinc finger, C2H2 type	3.5e-68	239.9	10	189-212:240-262:268-290:296-319:325-347:353-375:382-404:909-931:937-960:966-988
1108	SET	SET domain	0.0012	-18.9	1	37-175
1108	zf-BED	BED zinc finger	0.1	10.4	2	276-320:922-961
1108	FYVE	FYVE zinc finger	6.8	-24.6	1	262-364
1109	Nucleosid e_tra2	Na+ dependent nucleoside transporter	2.5e-187	635.7	1	198-613
1109	TLC	TLC ATP/ADP transporter	5.6	-382.0	1	95-407
1109	ATP-synt_A	ATP synthase A chain	6.5	-79.5	1	365-503
1110	PHD	PHD-finger	5.2	-16.5	1	77-120
1113	Peptidase C48	Ulp1 protease family, C-terminal cataly	6.5e-34	126.1	1	254-415
1114	ras	Ras family	5.5e-31	116.4	1	54-222
1114	arf	ADP-ribosylation factor family	0.0054	-80.3	1	37-213
1115	SPRY	SPRY domain	7.3e-10	46.2	1	281-419
1115	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.00076	26.2	1	88-134
1116	LRR	Leucine Rich Repeat	2.1e-05	31.4	6	37-64:65-92:93-120:121-142:175-198:255-278
1117	RasGAP	GTPase-activator protein for Ras-like GTPase	6.4e-28	106.2	1	268-440
1117	PH	PH domain	0.28	15.7	1	7-78
1117	C2	C2 domain	1.9	-9.5	1	91-171
1117	bZIP	bZIP transcription factor	9	-6.7	1	997-1054
1118	COX3	Cytochrome c oxidase subunit III	1.7	-228.9	1	77-195
1118	sugar_tr	Sugar (and other) transporter	2.2	-179.3	1	32-413
1120	LMWPc	Low molecular weight phosphotyrosine protein	9.7e-56	198.6	1	7-138
1122	M	M protein repeat	6.6	13.1	2	148-168:216-236
1123	lectin_c	Lectin C-type domain	4.3e-11	50.3	1	579-646
1123	Tropomyosin	Tropomyosin	0.17	-90.1	1	304-500
1123	filament	Intermediate filament protein	0.17	-164.6	1	287-537
1123	spectrin	Spectrin repeat	5.8	-18.2	1	422-523
1127	vwa	von Willebrand factor type A domain	4.1e-63	223.1	1	252-450
1127	trypsin	Trypsin	2.1e-43	157.6	1	463-734

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1127	sushi	Sushi domain (SCR repeat)	2.4e-24	94.3	3	37-85:86-140:147-200
1128	Neur_chanmemb	Neurotransmitter-gated ion-channel tra	1e-88	308.2	1	84-334
1128	oxidored_q1_N	NADH-Ubiquinone oxidoreductase	7	-14.7	1	123-183
1129	C4	C-terminal tandem repeated domain in type 4	1.4e-148	507.0	2	477-584:585-699
1129	Collagen	Collagen triple helix repeat (20 copies)	1.9e-60	214.3	7	20-78:84-142:143-202:205-265:266-325:329-388:405-464
1132	filament	Intermediate filament protein	2.1	-193.2	1	90-330
1132	Tropomyosin	Tropomyosin	7.7	-119.0	1	151-353
1135	zf-C2H2	Zinc finger, C2H2 type	8.3e-26	99.2	5	278-303:312-339:345-369:375-399:405-429
1136	Ribosomal S2	Ribosomal protein S2	1.7e-78	274.2	1	34-198
1137	ATP-synt_ab_N	ATP synthase alpha/beta family, beta-ba	5.2e-26	99.8	1	63-129
1139	ATP-synt_ab	ATP synthase alpha/beta family, nucleot	3.1e-06	-33.1	1	71-183
1139	ATP-synt_ab_N	ATP synthase alpha/beta family, beta-ba	0.015	14.4	1	10-68
1140	ATP-synt_ab_N	ATP synthase alpha/beta family, beta-ba	5.2e-26	99.8	1	63-129
1140	ATP-synt_ab	ATP synthase alpha/beta family, nucleot	2.6e-06	-31.8	1	132-261
1141	ank	Ankyrin repeat	6.7e-34	126.0	3	463-495:496-528:529-561
1141	BRCT	BRCA1 C Terminus (BRCT) domain	1.5e-15	65.1	2	578-689:705-812
1141	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.083	9.7	1	86-122
1143	GTP_EFTU	Elongation factor Tu GTP binding domain	0.031	-59.5	1	68-140
1145	RF-1	Peptidyl-tRNA hydrolase domain	3.3e-05	10.5	1	46-159
1146	RF-1	Peptidyl-tRNA hydrolase domain	6.4	-51.5	1	46-114
1148	WD40	WD domain, G-beta repeat	6.9e-07	36.3	1	44-80
1149	Band 41	FERM domain (Band 4.1 family)	1.1e-77	271.5	1	45-235
1150	Metallothionein	Plant PEC family metallothionein	5	-38.3	1	90-139
1153	pkinase	Protein kinase domain	1.7e-90	314.1	1	43-299

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1155	SCP	SCP-like extracellular protein	6.2e-14	56.1	1	28-200
1156	ras	Ras family	1.5e-15	40.0	1	5-98
1159	6PF2K	6-phosphofructo-2-kinase	5.2e-152	518.4	1	26-249
1159	PGAM	Phosphoglycerate mutase family	6e-100	345.5	1	250-435
1160	zf-C2H2	Zinc finger, C2H2 type	1.3e-104	361.0	13	223-245;251-273;279-301;307-329;335-357;363-385;391-413;419-441;447-469;475-497;501-523;529-551;557-579
1160	PHD	PHD-finger	0.55	-7.5	1	448-509
1160	zf-BED	BED zinc finger	1.4	0.3	2	264-302;542-580
1160	DnaJ_CX XCXGX G	DnaJ central domain (4 repeats)	2.9	-45.3	1	509-572
1160	LIM	LIM domain	7.7	-19.2	1	477-539
1160	TFIIIS	Transcription factor S-II (TFIIIS)	7.9	-6.3	1	450-485
1162	Patatin	Patatin-like phospholipase	0.00033	-0.6	1	1-171
1163	pkinase	Protein kinase domain	2.3e-94	326.9	1	53-303
1163	RIO1	RIO1/ZK632.3/MJ0444 family	0.37	-100.2	1	47-245
1164	Oxysterol BP	Oxysterol-binding protein	3.8e-47	170.0	1	173-571
1166	OATP_C	Organic Anion Transporter Polypeptide (OATP)	9.9e-195	660.3	1	68-443
1166	OATP_N	Organic Anion Transporter Polypeptide (OATP)	1e-57	238.4	1	520-680
1166	7tm_5	7TM chemoreceptor	6.1	-167.4	1	184-503
1166	sugar tr	Sugar (and other) transporter	7.8	-195.1	1	48-570
1167	pentaxin	Pentaxin family	2.3e-07	-7.3	1	25-98
1168	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	1.7e-05	31.7	1	537-574
1169	Peptidase M1	Peptidase family M1	4e-156	532.1	1	69-458
1170	ig	Immunoglobulin domain	0.0016	25.1	1	30-109
1172	BTB	BTB/POZ domain	8.7e-33	122.4	1	104-214
1173	F-box	F-box domain	0.0019	24.9	1	16-64
1174	TPR	TPR Domain	4.5	9.3	1	301-334
1175	2OG- FeII_Oxy	2OG-Fe(II) oxygenase superfamily	1.6e-06	35.1	1	527-648
1175	TPR	TPR Domain	4.5	9.3	1	301-334
1176	2OG- FeII_Oxy	2OG-Fe(II) oxygenase superfamily	1.6e-06	35.1	1	557-678
1177	Na_Ca_E x	Sodium/calcium exchanger protein	1.1e-23	92.1	1	236-381
1179	lactamase B	Metallo-beta-lactamase superfamily	0.059	-4.5	1	13-212
1180	fibrinogen C	Fibrinogen beta and gamma chains, C-term	1.6e-32	121.5	1	207-416

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1181	MIF	Macrophage migration inhibitory factor (MIF)	7.6e-67	235.5	1	2-115
1182	SSF	Sodium:solute symporter family	1.7e-234	792.4	1	69-503
1182	xan_ur_pe rmease	Permease family	4.3	-193.8	1	143-514
1182	PNTB	NAD(P) transhydrogenase beta subunit	5	-389.7	1	33-325
1182	60KD_IM P	60Kd inner membrane protein	6.2	-129.6	1	113-241
1182	oxidored_ gl	NADH-Ubiquinone/plastoquinone	7.4	-169.1	1	87-298
1183	Cation_ef flux	Cation efflux family	2.3e-58	207.3	1	73-311
1185	AAA	ATPase family associated with various cel	4e-89	309.5	2	236-421:500-620
1185	HypB_Ur eG	HypB/UreG nucleotide-binding domain	3.2	-65.8	1	234-337
1185	ArsA_AT Pase	Anion-transporting ATPase	8.5	-195.7	1	234-482
1186	HCO3_co transp	HCO3- transporter family	0	1389.6	1	141-1023
1186	xan_ur_pe rmease	Permease family	0.33	-164.5	1	518-985
1187	homeobox	Homeobox domain	2.4e-16	67.7	1	51-107
1188	efhand	EF hand	6.7	9.0	1	13-41
1191	GST_C	Glutathione S-transferase, C-terminal domain	0.93	0.3	1	134-326
1194	PPR	PPR repeat	0.0019	24.9	1	14-48
1195	thioredo	Thioredoxin	0.018	-13.7	1	390-497
1197	ENV_pol yprotein	ENV polyprotein (coat polyprotein)	1.2e-08	-24.1	1	86-529
1200	UQ_con	Ubiquitin-conjugating enzyme	1.3e-23	91.9	1	60-190
1202	7tm_1	7 transmembrane receptor (rhodopsin family)	2.1e-23	91.2	1	59-306
1202	7tm_5	7TM chemoreceptor	4.3	-164.7	1	37-314
1203	7tm_1	7 transmembrane receptor (rhodopsin family)	5.9e-37	136.2	1	59-341
1203	7tm_5	7TM chemoreceptor	2.2	-159.4	1	37-338
1204	SH3	SH3 domain	2.5e-05	31.1	1	257-317
1204	UBA	UBA/TS-N domain	0.00013	28.8	1	36-76
1204	PGAM	Phosphoglycerate mutase family	0.00044	-75.5	1	438-625
1205	heme_1	Heme/Steroid binding domain	0.00053	19.5	1	37-112
1207	transmem brane4	Tetraspanin family	0.29	-69.8	1	11-110
1208	OATP_C	Organic Anion Transporter Polypeptide	1.3e-135	464.0	1	148-524
1208	kazal	Kazal-type serine protease inhibitor d	0.11	4.9	1	555-601
1208	sugar_tr	Sugar (and other) transporter	0.15	-145.7	1	128-626
1208	lig_chan	Ligand-gated ion channel	3.3	-153.6	1	193-524
1208	7tm_1	7 transmembrane receptor	4.2	-115.5	1	177-473
1208	Cytidylyl rans	Phosphatidate cytidylyltransferase	4.6	-87.5	1	177-268
1215	thyroglob ulin_1	Thyroglobulin type-1 repeat	3.2e-35	130.4	2	90-153:216-281
1215	kazal	Kazal-type serine protease inhibitor	7.5e-09	42.8	1	40-84
1215	efhand	EF hand	0.057	20.0	2	351-

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						379:388-416
1216	annexin	Annexin	7e-61	215.7	2	72-139:144-211
1217	annexin	Annexin	2.2e-86	300.4	3	44-111:116-183:199-267
1218	annexin	Annexin	7e-61	215.7	2	44-111:116-183
1219	Armadillo_seg	Armadillo/beta-catenin-like repeat	5.3e-30	113.1	5	351-393:405-447:448-498:499-541:543-585
1219	GA	GA module	8.3	-5.7	1	180-224
1221	p450	Cytochrome P450	1e-122	421.1	1	30-483
1222	p450	Cytochrome P450	1.7e-06	-35.8	1	30-328
1223	A2M_N	Alpha-2-macroglobulin family N-terminal regi	4.3e-12	-71.6	1	1-468
1225	Thymosin	Thymosin beta-4 family	2.3e-16	67.8	1	2-41
1227	WD40	WD domain, G-beta repeat	1.6e-37	138.1	7	115-151:165-201:207-244:250-286:293-328:334-370:391-431
1236	F-box	F-box domain	5e-07	36.8	1	210-258
1236	UvrD-helicase	UvrD/REP helicase	0.00011	-157.2	1	441-920
1237	LRR	Leucine Rich Repeat	2.3e-25	97.7	8	42-65:66-88:89-111:112-134:135-157:158-180:181-203:204-227
1238	TPR	TPR Domain	9.6e-54	192.0	10	22-55:56-86:87-120:121-154:155-188:189-222:223-255:290-323:328-361:362-395
1241	cadherin	Cadherin domain	0.00011	29.0	2	48-151:165-254
1243	TTL	Tubulin-tyrosine ligase family	3.6e-31	117.0	1	1-225
1245	UCH-1	Ubiquitin carboxyl-terminal hydrolases famil	4.2e-08	40.4	1	190-221

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1245	zf-UBP	Zn-finger in ubiquitin-hydrolases and other	1.2e-05	32.2	1	62-148
1246	ank	Ankyrin repeat	3.8e-92	319.5	11	12-44:45-77:79-111:112-144:145-177:179-211:212-244:245-277:278-310:312-344:345-374
1250	C1q	C1q domain	0.00033	-3.8	1	827-946
1250	filament	Intermediate filament protein	1.3	-187.6	1	360-645
1250	spectrin	Spectrin repeat	1.9	-12.5	1	492-591
1250	TNF	TNF(Tumor Necrosis Factor) family	4.1	-20.5	1	835-946
1250	Apolipoprotein	Apolipoprotein A1/A4/E family	5.7	-113.3	1	200-469
1250	sigma70	Sigma-70 factor	9	-116.2	1	462-650
1252	laminin_Nterm	Laminin N-terminal (Domain VI)	3.2e-52	186.9	1	1-223
1252	laminin_EGF	Laminin EGF-like (Domains III and V)	2e-38	141.0	3	225-292:295-355:358-409
1252	NTR	NTR/C345C module	7.2e-30	112.6	1	479-591
1252	Keratin_B2	Keratin, high sulfur B2 protein	6.7	-81.2	1	318-451
1253	enolase	Enolase	0.038	-162.0	1	11-136
1256	HIT	HIT family	1.5e-55	198.0	1	51-162
1257	ank	Ankyrin repeat	5.9e-24	93.0	4	39-67:68-100:101-133:134-164
1258	lectin_c	Lectin C-type domain	1.3e-21	85.2	1	51-181
1258	lectin_c	Lectin C-type domain	1.3e-21	85.2	1	51-181
1262	mito_carr	Mitochondrial carrier protein	8.7e-67	235.3	3	7-100:102-221:224-319
1263	serpin	Serpin (serine protease inhibitor)	8.7e-109	374.8	1	87-463
1264	MHC_I	Class I Histocompatibility antigen, domains	1.3e-141	483.9	1	25-203
1265	MHC_I	Class I Histocompatibility antigen, domains	2.3e-130	446.5	1	25-203
1266	MHC_I	Class I Histocompatibility antigen, domains	2.7e-117	403.1	1	22-187
1267	MHC_I	Class I Histocompatibility antigen, domains	1.1e-115	397.7	1	25-196
1268	MHC_I	Class I Histocompatibility antigen, domains	2.9e-118	406.3	1	25-196
1268	ig	Immunoglobulin domain	9.5e-08	39.2	1	221-286
1269	MHC_I	Class I Histocompatibility antigen, domains	4.2e-132	452.3	1	25-204
1271	MHC_I	Class I Histocompatibility antigen, domains	1e-144	494.2	1	25-203
1272	MHC_I	Class I Histocompatibility antigen, domains	2.8e-95	329.9	1	25-204

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1273	MHC_I	domains Class I Histocompatibility antigen, domains	3.3e-140	479.2	1	25-203
1274	Kelch	Kelch motif	1.4e-91	317.7	6	271-316;318-366;368-413;415-460;462-507;509-554
1274	BTB	BTB/POZ domain	3.1e-38	140.4	1	23-130
1275	IQ	IQ calmodulin-binding motif	0.0037	23.9	1	394-414
1276	Glycos_tr ansf_2	Glycosyl transferase	2.2e-22	87.8	1	126-308
1276	Ricin_B_1 ectin	QXW lectin repeat	0.0045	23.7	2	478-518;520-557
1280	ig	Immunoglobulin domain	3.4e-05	30.7	2	62-145;174-240
1281	UCH-1	Ubiquitin carboxyl-terminal hydrolases fam1	1.5e-12	55.1	1	100-131
1285	UBX	UBX domain	2.9e-22	87.4	1	205-284
1286	UBX	UBX domain	2.9e-22	87.4	1	251-330
1287	DnaJ	DnaJ domain	9e-38	138.9	1	8-70
1288	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	7.9e-09	42.8	1	9-57
1288	zf-B_box	B-box zinc finger	1.8e-06	34.9	1	177-217
1288	SPRY	SPRY domain	0.0016	14.8	1	479-605
1288	zf-UBR1	Putative zinc finger in N-recogin	2.2	-18.4	1	180-237
1289	PKD	PKD domain	0	1026.8	16	293-372;397-483;485-568;572-677;679-763;765-846;849-931;933-1017;1019-1099;1101-1183;1185-1269;1271-1353;1355-1438;1440-1522;1524-1612;1614-1696
1289	REJ	REJ domain	2.3e-290	978.0	1	1723-2248
1289	PLAT	PLAT/LH2 domain	2.6e-25	97.5	1	2673-2789
1289	GPS	Latrophilin/CL-1-like GPS domain	1.1e-15	65.5	1	2566-2615
1289	lectin_c	Lectin C-type domain	0.59	-23.7	1	2-87
1289	DUF26	Domain of unknown function DUF26	7.8	-16.5	1	2220-2262
1290	CNH	CNH domain	2.7e-24	94.2	1	69-375
1292	RhoGAP	RhoGAP domain	1.8e-59	211.0	1	125-279
1293	Peptidase M10	Matrixin	8.2e-110	378.2	1	48-211
1293	fn2	Fibronectin type II domain	3.5e-84	293.1	3	230-271;288-

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						329:347-388
1293	hemopexin	Hemopexin	1e-33	125.5	3	486-530:608-654:656-699
1294	Peptidase M10	Matrixin	8.2e-110	378.2	1	48-211
1294	fn2	Fibronectin type II domain	3.5e-84	293.1	3	230-271:288-329:347-388
1297	UQ_con	Ubiquitin-conjugating enzyme	3.3e-20	80.6	1	400-557
1297	TT_ORF2	TT viral ORF2	3.6	-92.9	1	546-667
1297	UQ_con	Ubiquitin-conjugating enzyme	3.3e-20	80.6	1	400-557
1297	TT_ORF2	TT viral ORF2	3.6	-92.9	1	546-667
1298	ig	Immunoglobulin domain	2.9e-236	798.3	21	25-84:119-177:245-303:339-397:432-490:524-583:618-676:709-768:801-859:894-954:990-1048:1121-1179:1217-1275:1382-1440:1477-1535:1569-1627:1664-1722:1756-1816:1851-1911:1947-2005:2040-2098
1298	Adeno_E3_CR1	Adenovirus E3 region protein CR1	0.062	-3.7	1	1212-1288
1299	cNMP_binding	Cyclic nucleotide-binding domain	6.2e-28	106.2	1	363-459
1299	ion_trans	Ion transport protein	8.9e-21	82.5	1	69-265
1299	ATP-synt_DE	ATP synthase, Delta/Epsilon chain, long	6.8	6.0	1	478-525
1301	RGS	Regulator of G protein signaling domain	1.6e-49	177.9	1	56-172
1302	THF_DH_G_CYH_C	Tetrahydrofolate dehydrogenase/cyclohyd	1e-99	344.7	1	60-235
1303	gln-synt	Glutamine synthetase	1.3e-179	610.1	1	1-321
1304	gln-synt	Glutamine synthetase	0.13	-214.4	1	1-95
1305	SCAN	SCAN domain	7.5e-61	215.6	1	42-137
1306	Methyltransf_5	MraW methylase family	4.6e-06	-96.4	1	70-205
1307	ion_trans	Ion transport protein	8.3e-18	72.6	1	438-672
1307	ank	Ankyrin repeat	1.9e-06	34.8	2	77-108:163-

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1307	Srg	C.elegans Srg family integral membrane prot	8.1	-222.4	1	195 418-669
1310	Ammoniu m_transp	Ammonium Transporter Family	1.9e-56	200.9	1	25-429
1310	FecCD	FecCD transport family	0.89	-200.6	1	97-331
1311	Kelch	Kelch motif	2.6e-60	213.8	6	311- 359:361- 411:413- 458:460- 505:507- 556:559- 606
1311	BTB	BTB/POZ domain	1.6e-28	108.1	1	23-181
1313	zf-B_box	B-box zinc finger	9.6e-30	112.2	2	92- 133:418- 459
1313	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	1.3e-23	91.8	2	15-59:341- 385
1313	SPRY	SPRY domain	3.2e-23	90.6	1	672-813
1313	PHD	PHD-finger	0.97	-9.8	1	14-62
1313	NB-ARC	NB-ARC domain	9.5	-151.3	1	48-311
1313	zf-UBR1	Putative zinc finger in N-recogrin	9.8	-24.7	1	421-470
1314	SRCR	Scavenger receptor cysteine-rich domain	1.7e-25	98.1	1	37-133
1315	adh_short	short chain dehydrogenase	5.8e-33	122.9	1	38-293
1317	ANP	Atrial natriuretic peptide	1.2e-51	185.0	1	43-150
1318	BGF	BGF-like domain	3.6e-26	100.4	6	512- 547:553- 591:597- 634:640- 675:747- 781:787- 827
1318	ig	Immunoglobulin domain	2.1e-06	34.7	1	170-227
1318	TIL	Trypsin inhibitor like cysteine rich dom	1.9	-6.8	1	741-787
1318	ldl_recept _a	Low-density lipoprotein receptor domain	6.6	-8.8	1	551-593
1318	TILa	TILa domain	6.9	-8.5	1	733-792
1321	FKBP	FKBP-type peptidyl-prolyl cis-trans isomeras	5.7e-45	162.8	1	234-336
1322	PX	PX domain	1.1e-25	98.8	1	273-382
1322	SH3	SH3 domain	6.5e-12	53.0	1	3-59
1323	ras	Ras family	5.2e-16	45.4	1	35-218
1323	GTP_EFT U	Elongation factor Tu GTP binding domain	0.53	-76.1	1	31-223
1323	MobB	Molybdopterin guanine dinucleotide synthesis	4.8	-45.8	1	36-150
1323	GTP_CD C	Cell division protein	5.7	-217.6	1	36-247
1324	pkinase	Protein kinase domain	4.6e-51	183.1	1	199-527
1325	pkinase	Protein kinase domain	4.6e-51	183.1	1	199-527
1327	Peptidase _C1	Papain family cysteine protease	7.3e- 110	378.4	1	73-349
1328	Peptidase _C1	Papain family cysteine protease	7.3e- 110	378.4	1	114-390
1330	ig	Immunoglobulin domain	4.6e-07	36.9	2	52- 108:145-

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1332	Sec7	Sec7 domain	1.3e-71	251.3	1	201
1332	PH	PH domain	0.25	16.2	1	345-536
1332	IQ	IQ calmodulin-binding motif	0.35	17.4	1	567-676
1333	cadherin	Cadherin domain	1.6e-94	327.4	5	13-33
						68-159:173-268:282-386:399-490:503-600
1333	Cadherin_C term	Cadherin cytoplasmic region	1.6e-76	267.6	1	648-819
1335	lipocalin	Lipocalin / cytosolic fatty-acid binding pr	7.5e-35	129.2	1	5-133
1336	SH3	SH3 domain	2.4e-10	47.8	1	495-549
1336	PID	Phosphotyrosine interaction domain (PTB/PID)	0.29	-13.3	1	52-181
1336	SAM_PNT	Sterile alpha motif (SAM)/Pointed domain	6	-25.0	1	608-687
1336	SH2	SH2 domain	8.6	-25.5	1	268-347
1338	UCH-2	Ubiquitin carboxyl-terminal hydrolase	1.5e-20	81.7	1	350-411
1338	UCH-1	Ubiquitin carboxyl-terminal hydrolases	1.1e-12	55.5	1	111-142
1338	TFIIA	Transcription factor IIA, alpha/beta s	1.7	-151.4	1	386-698
1338	Tymo_45 kd_70kd	Tymovirus 45/70Kd protein	3.3	-291.1	1	310-647
1338	MARCKS	MARCKS family	6.8	-95.6	1	573-865
1340	RCC1	Regulator of chromosome condensation (RCC1)	5.6e-10	46.6	6	77-136:140-192:195-245:248-298:301-356:359-406
1342	Peptidase_M1	Peptidase family M1	4.6	-194.5	1	159-470
1342	ubiquitin	Ubiquitin family	5.4	-4.4	1	538-616
1345	crystall	Beta/Gamma crystallin	3.3e-38	140.4	2	1-53:61-144
1345	WD40	WD domain, G-beta repeat	6.7e-09	43.0	1	269-305
1348	AAA	ATPase family associated with various cellu	0.97	-33.9	1	131-307
1349	AAA	ATPase family associated with various cellu	0.97	-33.9	1	131-307
1351	PHD	PHD-finger	6.3e-15	63.0	1	190-237
1351	PHD	PHD-finger	6.3e-15	63.0	1	190-237
1353	CUB	CUB domain	3.3e-13	57.3	1	416-524
1353	sushi	Sushi domain (SCR repeat)	1.5e-05	31.9	1	357-412
1355	SPRY	SPRY domain	3.4e-19	77.2	1	396-519
1355	fn3	Fibronectin type III domain	1e-08	42.4	1	259-345
1355	zf-B_box	B-box zinc finger	8.9e-07	35.9	1	44-86
1356	Patatin	Patatin-like phospholipase	0.59	-42.1	1	10-118
1358	ELM2	ELM2 domain	3.2e-21	84.0	1	195-256
1358	myb_DNA-binding	Myb-like DNA-binding domain	1.1e-09	45.6	1	299-345
1359	Reprolysin	Reprolysin (M12B) family zinc metallo	1.4e-12	55.3	1	239-457
1359	Pep_M12B_propep	Reprolysin family propeptide	3.6e-12	53.9	1	90-216

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1359	tsp_1	Thrombospondin type I domain	3.4e-07	37.4	5	551-601:829-884:888-944:946-1002:1007-1057
1359	EB	EB module	8.4	-11.7	1	452-531
1363	ank	Ankyrin repeat	6.2e-26	99.6	5	72-104:105-137:138-170:180-212:222-255
1366	C2	C2 domain	1e-75	264.9	2	161-247:293-381
1367	C2	C2 domain	1e-75	264.9	2	161-247:293-381
1368	p450	Cytochrome P450	5e-110	378.9	1	47-502
1370	aa_permeases	Amino acid permease	1.3e-08	-108.1	1	49-452
1370	Neur_chan_memb	Neurotransmitter-gated ion-channel tr	0.37	-96.5	1	330-585
1370	ion_trans	Ion transport protein	1.4	-2.3	1	288-615
1370	Transp_cyt_pur	Permease for cytosine/purines, uracil	1.5	-178.6	1	47-442
1370	Aa_trans	Transmembrane amino acid transporter	1.9	-190.3	1	67-409
1370	DUF140	Domain of unknown function DUF140	2.6	-156.6	1	109-312
1370	Nucleoside_trans	Nucleoside transporter	3.2	-154.9	1	357-658
1370	xan_ur_permease	Permease family	4.1	-193.2	1	56-429
1370	DUF6	Integral membrane protein DUF6	7.1	-22.4	1	536-671
1370	NADHdh	NADH dehydrogenase	7.3	-213.9	1	212-616
1370	SNF	Sodium:neurotransmitter symporter fam	9.2	-458.6	1	117-450
1372	kinesin	Kinesin motor domain	2.1e-135	463.2	1	53-341
1372	Translin	Translin family	10	-82.4	1	315-462
1373	LRR	Leucine Rich Repeat	1.9e-27	104.6	7	60-83:84-107:108-131:132-155:157-180:181-204:205-225
1373	ig	Immunoglobulin domain	1.2e-05	32.1	1	310-368
1373	LRRCT	Leucine rich repeat C-terminal domain	4.7e-05	30.2	1	249-294
1373	fn3	Fibronectin type III domain	7.1e-05	29.6	1	425-505
1373	LRNT	Leucine rich repeat N-terminal domain	0.83	13.7	1	27-58
1374	kinesin	Kinesin motor domain	9.5e-14	1.2	1	1-131
1375	zf-DHHC	DHHC zinc finger domain	9.1e-34	125.6	1	101-165
1376	EGF	EGF-like domain	3.2e-45	163.6	9	49-84:90-126:132-167:177-213:217-252:286-

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						321:327-362:368-401:407-442
1376	CUB	CUB domain	9.6e-18	72.4	1	809-918
1376	TIL	Trypsin Inhibitor like cysteine rich domain	0.73	-2.0	1	84-132
1376	Keratin_B2	Keratin, high sulfur B2 protein	0.9	-67.6	1	111-242
1376	granulin	Granulin	1.1	-12.1	1	285-323
1376	metallothionein	Metallothionein	5.9	-10.3	1	363-431
1376	DUF141	Domain of unknown function DUF141	6.7	-15.1	1	799-922
1380	ion transport protein	Ion transport protein	0.066	16.8	1	153-318
1380	ABC2_membrane	ABC-2 type transporter	6.1	-130.1	1	145-334
1380	oxidoreductase	NADH-Ubiquinone/plastoquinone	6.2	-167.5	1	46-317
1380	OATP_C	Organic Anion Transporter Polypeptide	9.5	-236.4	1	75-316
1381	kinase	Protein kinase domain	9.7e-80	278.3	1	205-486
1384	Ig	Immunoglobulin domain	0.00034	27.4	1	65-142
1388	LRR	Leucine Rich Repeat	4.9e-16	66.7	4	20-43:44-67:68-91:92-115
1388	LRRCT	Leucine rich repeat C-terminal domain	7.4e-09	42.9	1	125-175
1388	GPS	Latrophilin/CL-1-like GPS domain	0.0041	20.8	1	641-693
1388	HRM	Hormone receptor domain	0.0076	16.3	1	285-354
1388	7tm_2	7 transmembrane receptor (Secretin family)	0.01	-96.0	1	704-981
1388	Ig	Immunoglobulin domain	3.3	10.9	1	196-265
1389	MACPF	MAC/Perforin domain	0.016	-71.3	1	30-313
1391	HD	HD domain	8.3e-07	36.1	1	32-127
1392	efhand	EF hand	1.5e-05	31.8	2	1-25:33-61
1394	MORN	MORN repeat	1.1e-32	122.1	7	39-61:62-85:86-108:109-131:132-154:155-177:178-200
1395	MORN	MORN repeat	3.1e-31	117.2	6	39-61:62-85:86-108:143-165:166-188:189-211
1396	EPH_lbd	Ephrin receptor ligand binding domain	4.6e-135	462.1	1	31-204
1396	kinase	Protein kinase domain	8e-74	258.7	1	635-892
1396	fn3	Fibronectin type III domain	2.5e-31	117.5	2	329-425:437-524
1396	SAM	SAM domain (Steric alpha motif)	2.5e-21	84.3	1	928-992
1396	PHD	PHD-finger	3.7	-15.1	1	259-296
1397	Exo_endo_phos	Endonuclease/Exonuclease/phosphatase	0.34	-11.2	1	357-643
1397	sugar_tr	Sugar (and other) transporter	8.8	-196.7	1	7-343
1397	Exo_endo_phos	Endonuclease/Exonuclease/phosphatase	0.34	-11.2	1	357-643

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1397	sugar_tr	Sugar (and other) transporter	8.8	-196.7	1	7-343
1401	homeobox	Homeobox domain	0.097	-3.3	1	78-126
1403	Tropomyosin	Tropomyosin	2.4	-110.1	1	341-547
1403	filament	Intermediate filament protein	2.6	-195.5	1	412-760
1403	UvrD-helicase	UvrD/REP helicase	8	-249.4	1	153-720
1403	Myosin_tail	Myosin tail	8.9	-551.2	1	20-777
1404	Clq	Clq domain	0.00033	-3.8	1	827-946
1404	filament	Intermediate filament protein	1.3	-187.6	1	360-645
1404	spectrin	Spectrin repeat	1.9	-12.5	1	492-591
1404	TNF	TNF(Tumor Necrosis Factor) family	4.1	-20.5	1	835-946
1404	Apolipoprotein	Apolipoprotein A I/A4/E family	5.7	-113.3	1	200-469
1404	sigma70	Sigma-70 factor	9	-116.2	1	462-650
1405	Cache	Cache domain	6.5e-12	53.0	2	402-481:721-813
1406	ASC	Amiloride-sensitive sodium channel	2e-125	430.1	1	159-579
1407	pkinase	Protein kinase domain	0.22	-115.5	1	5-217
1408	PBP	Phosphatidylethanolamine-binding protein	1.8e-71	250.9	1	1-167
1410	abhydrolase	alpha/beta hydrolase fold	1.4	-12.1	1	75-318
1412	rrm	RNA recognition motif.	2.7e-12	54.3	3	259-329:360-433:477-550
1414	DEP	Domain found in Dishevelled, Egl-10, and Ple	3.5e-32	120.3	2	173-247:275-349
1414	PH	PH domain	1e-09	45.7	1	29-144
1414	PDZ	PDZ domain (Also known as DHR or GLGF)	0.013	18.7	2	375-456:460-531
1418	SCAN	SCAN domain	2.6e-51	183.9	1	36-131
1418	zf-C2H2	Zinc finger, C2H2 type	4.6e-25	96.7	5	406-428:435-457:463-485:522-545:553-575
1418	KRAB	KRAB box	0.023	14.2	1	220-260
1418	zf-BED	BED zinc finger	10	-7.4	1	438-486
1419	EGF	EGF-like domain	5.1e-28	106.5	9	13-40:45-72:77-104:109-136:141-168:173-200:205-232:237-264:269-296
1419	Keratin_B2	Keratin, high sulfur B2 protein	1.2	-69.3	1	100-251
1419	Gamma-thionin	Gamma-thionins family	4.8	-8.7	1	124-163

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1421	Ferric_red uct	Ferric reductase like transmembrane com	2.8e-64	227.0	1	63-564
1421	NAD_bin ding	Oxidoreductase NAD-binding domain	4	-34.1	1	381-551
1421	FAD_bin ding_6	Oxidoreductase FAD-binding domain	5	-28.7	1	245-335
1421	rubredoxi n	Rubredoxin	6.9	-11.0	1	409-436
1422	zf-C2H2	Zinc finger, C2H2 type	5e-12	53.4	2	1057- 1079:1085- 1109
1422	zf-TRAF	TRAF-type zinc finger	5	-6.6	1	1056-1101
1422	Tymo_45 kd_70kd	Tymovirus 45/70Kd protein	6	-298.2	1	343-700
1422	zf-BED	BED zinc finger	7.1	-6.0	1	1070-1110
1423	7tm_5	7TM chemoreceptor	3.3	-162.5	1	181-451
1423	Cytidylt rans	Phosphatidate cytidyltransferase	5	-87.9	1	21-135
1424	CAP_GL Y	CAP-Gly domain	7.2e-46	165.8	2	196- 238:398- 440
1424	ank	Ankyrin repeat	9.2e-09	42.5	3	1-40:42- 76:79-111
1425	PAP2	PAP2 superfamily	1.5e-08	41.8	1	166-313
1426	SCAN	SCAN domain	6.2e-70	245.7	1	33-128
1426	zf-C2H2	Zinc finger, C2H2 type	1.5e-44	161.4	6	239- 261:267- 289:295- 317:323- 345:351- 373:379- 401
1426	zf-BED	BED zinc finger	0.67	3.1	1	280-318
1426	DC1	DC1 domain	5.1	3.6	1	295-338
1426	zf-C4	Zinc finger, C4 type (two domains)	9.7	-55.1	1	323-364
1427	xan_ur_pe rmease	Permease family	7.1	-199.5	1	104-453
1428	LRR	Leucine Rich Repeat	2.3e-16	67.8	4	80- 103:104- 127:128- 151:152- 175
1428	LRRCT	Leucine rich repeat C-terminal domain	0.00079	26.2	1	185-234
1431	PH	PH domain	7.6e-15	62.8	1	19-117
1432	PH	PH domain	6.4e-21	82.9	1	55-153
1434	filament	Intermediate filament protein	2.9	-196.8	1	128-488
1434	K-box	K-box region	4.1	-38.4	1	277-357
1434	OspD	Borrelia outer surface protein D	5.3	-69.3	1	151-409
1434	Apolipop rotein	Apolipoprotein A1/A4/E family	5.7	-113.3	1	56-318
1437	RhoGAP	RhoGAP domain	5.7e-57	202.7	1	1152-1305
1437	PH	PH domain	3.5e-18	73.8	1	922-1030
1437	PDZ	PDZ domain (Also known as DHR or GLGF)	0.0013	25.4	1	28-128
1441	annexin	Annexin	4.1e- 109	375.9	4	18-79:80- 135:151- 219:227- 294

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1443	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.0075	18.7	1	1256-1307
1443	PHD	PHD-finger	0.18	-2.9	1	1255-1310
1443	WD40	WD domain, G-beta repeat	2.9	13.4	1	189-225
1443	K_tetra	K+ channel tetramerisation domain	3.5	-39.5	1	830-916
1443	Clathrin	Region in Clathrin and VPS	8.9	-25.2	1	976-1129
1444	R1fa	Regulatory subunit of type II PKA R-s	0.31	10.8	1	17-54
1444	Alpha_adaptin_C	Alpha adaptin AP2, C-terminal domain	7	-63.6	1	91-155
1445	TK	Thymidine kinase	3.4e-98	339.6	1	61-231
1446	GDPD	Glycerophosphoryl diester phosphodiester	5.4e-08	40.0	1	154-403
1446	Glycosyltransferase	Glycosyl transferase	4.9	-87.4	1	22-145
1447	adh_short	short chain dehydrogenase	0.98	-91.8	1	2-196
1448	K-box	K-box region	0.64	-28.5	1	61-126
1449	PHD	PHD-finger	0.01	8.4	1	1-42
1452	rrm	RNA recognition motif	4.5e-19	76.8	1	77-148
1454	rvt	Reverse transcriptase	3.6e-34	126.9	1	385-570
1454	Gag_MA	Matrix protein (MA), p15	0.0018	-21.0	1	10-131
1454	Gag_p30	Gag P30 core shell protein	0.54	-80.3	1	211-390
1458	COX5A	Cytochrome c oxidase subunit Va	1.2e-55	198.3	1	42-131
1459	Guanylate_kin	Guanylate kinase	6.2e-38	139.4	1	515-624
1459	PDZ	PDZ domain (Also known as DHR or GLGF)	6.8e-11	49.6	1	256-335
1459	SH3	SH3 domain	0.027	5.9	1	348-415
1459	L27	L27 domain	0.049	20.1	1	186-238
1459	Caulimo_mov	Caulimovirus movement protein	7.1	-185.3	1	420-673
1459	A_deaminase	Adenosine/AMP deaminase	7.8	-138.5	1	64-421
1461	hexokinase	Hexokinase	4.3e-284	957.2	1	53-499
1463	Occludin	Occludin/ELL family	6.3	-249.1	1	33-394
1464	trypsin	Trypsin	4.7e-72	252.8	1	30-232
1466	DDHD	DDHD domain	8.6e-117	401.4	1	613-860
1466	DUF203	Domain of unknown function	8.7	-69.8	1	254-460
1467	Glycosyltransferase	Glycosyl transferases group 1	1.8e-27	104.7	1	286-470
1468	EMP24_GP25L	emp24/gp25L/p24 family	3.5e-70	246.6	1	5-183
1469	EMP24_GP25L	emp24/gp25L/p24 family	3.5e-81	283.1	1	5-208
1470	14-3-3	14-3-3 protein	2.2e-142	486.5	1	5-249
1471	filament	Intermediate filament protein	0.53	-177.6	1	2-249
1471	spectrin	Spectrin repeat	7.1	-19.1	1	34-130
1472	Min3_slv	Min3/saliva family	5.4	-31.6	1	35-139
1472	ATP-synt_A	ATP synthase A chain	7.4	-80.2	1	91-214
1474	zf-CCCH	Zinc finger C-x8-C-x5-C-x3-H type	4.3e-07	37.0	2	531-557;663-688
1474	zf-C2H2	Zinc finger, C2H2 type	6.3	11.3	2	205-229;618-642
1475	SpoU_methylase	SpoU rRNA Methylase family	2.1e-27	104.5	1	145-301

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1476	filament	Intermediate filament protein	2.4	-194.5	1	427-720
1476	Tropomyosin	Tropomyosin	3	-111.7	1	539-761
1476	spectrin	Spectrin repeat	3.3	-15.3	1	427-526
1476	K-box	K-box region	7.1	-41.3	1	248-335
1476	Borrelia_orfA	Borrelia ORF-A	9.8	-102.9	1	440-736
1477	MMR_HS_R1	GTPase of unknown function	1.5e-90	314.2	1	178-521
1477	DUF258	Protein of unknown function, DUF258	9.6	-84.6	1	343-465
1479	RNase_PH	3' exoribonuclease family	1.2e-96	334.5	2	48-251:358-581
1479	S1	S1 RNA binding domain	0.057	19.9	1	675-750
1479	KH-domain	KH domain	0.35	9.3	1	609-651
1482	COLF1	Fibrillar collagen C-terminal domain	5.8e-29	109.6	1	97-277
1482	Collagen	Collagen triple helix repeat (20 copies)	8.5e-05	29.4	1	2-61
1483	COLF1	Fibrillar collagen C-terminal domain	1.6e-35	131.4	1	110-293
1483	Collagen	Collagen triple helix repeat (20 copies)	8.5e-05	29.4	1	2-61
1484	CH	Calponin homology (CH) domain	5.6e-14	39.9	1	4-104
1485	MoaE	MoaE protein	7	-55.7	1	21-96
1486	zf-C2H2	Zinc finger, C2H2 type	6.8	10.9	2	871-896:904-929
1487	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	2.6	-3.3	1	336-377
1489	FH2	Formin Homology 2 Domain	0.00017	-49.7	1	3-329
1490	AAA	ATPase family associated with various	4e-45	163.3	1	370-565
1490	SKI	Shikimate kinase	0.068	-52.5	1	369-506
1490	Viral_helicase1	Viral (Superfamily 1) RNA helicase	1.8	-67.0	1	371-563
1490	LON	ATP-dependent protease La (LON) domain	3.6	-69.4	1	12-220
1491	Tropomodulin	Tropomodulin	4.1e-78	272.9	1	34-402
1491	WH2	WH2 motif	0.83	16.1	1	534-533
1491	pkinase	Protein kinase domain	5.9	-136.0	1	334-538
1494	xan_ur_permease	Permease family	2.9	-189.3	1	72-377
1494	Na_sulph_symp	Sodium:sulfate symporter transmembran	5.3	-356.1	1	212-541
1494	Glycosyltransferase_4	Glycosyl transferase	7.3	-90.3	1	374-528
1494	STE3	Pheromone A receptor	7.5	-203.9	1	314-603
1494	DUF221	Domain of unknown function DUF221	9.6	-234.2	1	196-576
1494	7tm_5	7TM chemoreceptor	9.7	-171.0	1	122-365
1494	oxidored_q1	NADH-Ubiquinone/plastoquinone	9.7	-171.5	1	42-264
1495	lectin_c	Lectin C-type domain	4.7e-34	126.6	1	53-164
1496	Cytochrome b561	Cytochrome b561	2.1e-113	390.2	1	1-240
1498	Hydrolase	haloacid dehalogenase-like hydrolase	0.0045	16.9	1	31-443
1498	Cation_ATPase_C	Cation transporting ATPase, C-terminu	0.26	-25.8	1	535-706
1498	oxidored_q4	NADH-ubiquinone/plastoquinone oxidore	4.4	-34.0	1	631-705
1499	DEAD	DEAD/DEAH box helicase	3e-64	226.9	1	139-356
1499	helicase	Helicase conserved C-terminal domain	4.1e-32	120.1	1	447-518

SEQ ID	Model	Description	E-value	Score	Repeats	Position
	C					
1502	C2	C2 domain	1.3e-51	184.9	2	28-109:184-264
1503	ank	Ankyrin repeat	1.6e-31	118.2	6	743-775:776-807:808-840:842-874:875-915:916-948
1504	ank	Ankyrin repeat	1.4e-167	570.1	16	57-89:90-122:123-155:210-242:243-275:276-308:363-398:399-431:432-464:525-557:558-590:591-623:644-675:678-710:711-743:744-776
1504	SAM	SAM domain (Sterile alpha motif)	1.3e-09	45.4	1	872-934
1504	PARP	Poly(ADP-ribose) polymerase catalytic domain	0.022	-59.4	1	954-1161
1504	3Beta_HSD	3-beta hydroxysteroid dehydrogenase/isomerase	1.3	-204.0	1	208-536
1505	PHD	PHD-finger	0.26	-4.5	1	132-191
1505	DC1	DC1 domain	2.8	5.8	1	131-159
1506	PHD	PHD-finger	0.26	-4.5	1	156-215
1506	DC1	DC1 domain	2.8	5.8	1	155-183
1507	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	3.6e-08	40.6	1	224-261
1507	LON	ATP-dependent protease La (LON) domain	0.007	-21.6	1	305-510
1507	TPR	TPR Domain	1.6	13.6	2	41-74:75-108
1508	ig	Immunoglobulin domain	1e-76	268.3	12	78-131:171-245:276-330:364-432:463-516:552-623:654-705:740-797:828-880:914-981:1012-1067:1101-1169
1512	FYVE	FYVE zinc finger	3.2e-14	60.7	1	152-261
1512	HypA	Hydrogenase expression/synthesis hypA	0.81	-51.2	1	97-194
1512	RNA_PO L_M_15K	RNA polymerases M/15 Kd subunit	2.1	-1.7	1	160-212

SEQ ID	Model	Description	E-value	Score	Repeats	Position
	D					
1516	Gelsolin	Gelsolin repeat	0.76	5.9	1	1011-1052
1518	V1R	Vomeroneal organ pheromone receptor family	2.3e-11	41.1	1	73-338
1518	7tm_1	7 transmembrane receptor (rhodopsin family)	0.044	-69.7	1	62-325
1520	SPRY	SPRY domain	5.2e-24	93.2	1	95-230
1521	Tropomodulin	Tropomodulin	0.065	-128.4	1	214-495
1521	LRR	Leucine Rich Repeat	1.8	15.0	4	348-371:376-403:404-427:432-456
1522	zf-CCHC	Zinc knuckle	0.051	15.9	1	13-30
1523	Skp1	Skp1 family	6.3e-10	46.4	1	17-80
1524	RhoGAP	RhoGAP domain	4.2e-31	116.8	1	125-285
1525	UQ_con	Ubiquitin-conjugating enzyme	1.4e-39	144.9	1	1-126
1527	LRR	Leucine Rich Repeat	1.6e-35	131.4	9	86-109:110-133:134-157:158-181:182-205:206-229:230-251:254-277:279-302
1527	LRRNT	Leucine rich repeat N-terminal domain	6.6e-06	33.1	1	33-60
1527	LRRCT	Leucine rich repeat C-terminal domain	0.048	17.9	1	312-362
1528	K_tetra	K+ channel tetramerisation domain	0.0016	-5.0	1	117-220
1529	MORN	MORN repeat	1.8e-24	94.7	7	1049-1071:1072-1094:1100-1122:1123-1143:1151-1171:1198-1220:1221-1244
1529	VPS9	Vacuolar sorting protein 9 (VPS9) domain	8.3e-06	32.7	1	1551-1656
1529	RCC1	Regulator of chromosome condensation (RCC1)	8.5e-06	32.7	3	168-216:527-574:579-625
1529	RhoGEF	RhoGEF domain	0.097	-40.8	1	694-884
1529	PH	PH domain	0.23	16.5	1	901-1005
1530	profilin	Profilin	1.4e-63	224.6	1	3-135
1531	profilin	Profilin	7.4e-48	172.4	1	3-119
1532	60s_ribosomal	60s Acidic ribosomal protein	3.3	-34.2	1	39-153
1533	jmjC	jmjC domain	0.01	-0.1	1	1-50
1533	PHD	PHD-finger	0.042	2.9	2	508-549:609-655
1534	kinesin	Kinesin motor domain	3.5e-64	226.7	1	1-177
1536	aminotran_3	Aminotransferase class-III	7.8e-42	152.4	1	1-373

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1536	LEA	Late embryogenesis abundant protein	5.6	-4.7	1	109-177
1539	Hydrolase	haloacid dehalogenase-like hydrolase	0.0064	15.3	1	264-685
1539	Cation_A TPase_C	Cation transporting ATPase, C-terminu	4.6	-46.1	1	784-916
1541	TPR	TPR Domain	0.00036	27.3	2	135- 168:204- 237
1542	PCMT	Protein-L-isocaspate(D-aspartate) O- methyl	1.2e-11	21.8	1	9-224
1543	Peptidase C54	Peptidase family C54	3.1e-58	206.9	1	76-364
1545	homeobox	Homeobox domain	4.8e-26	100.0	1	233-286
1546	zf-C2H2	Zinc finger, C2H2 type	3.1e-85	296.6	14	14-36:42- 64:70- 92:99- 122:128- 150:163- 185:199- 221:227- 249:255- 277:283- 305:311- 333:339- 361:367- 389:395- 417
1546	TFIIIS	Transcription factor S-II (TFIIIS)	1.9	-1.0	1	202-237
1546	zF-BED	BED zinc finger	2.3	-1.7	1	324-362
1546	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	4.6	-5.4	1	341-372
1547	Ribosoma l_S5	Ribosomal protein S5	3.4e-09	44.0	1	222-352
1548	IQ	IQ calmodulin-binding motif	1e-24	95.5	5	748- 768:771- 791:794- 814:935- 955:958- 978
1552	DUF6	Integral membrane protein DUF6	0.14	6.6	1	150-279
1552	SBF	Sodium Bile acid symporter family	9.2	-75.2	1	143-321
1553	zf-C2H2	Zinc finger, C2H2 type	2.1e-05	31.4	3	80- 105:107- 130:144- 169
1554	F-box	F-box domain	7.7e-05	29.5	1	4-52
1555	Ran BP1	RanBP1 domain	1.1e-88	308.0	1	37-161
1555	WH1	WH1 domain	6.8	-26.8	1	45-159
1556	actin	Actin	8.4e- 151	514.4	1	1-372
1557	GTP_EFT U	Elongation factor Tu GTP binding doma	9.7	-93.3	1	91-355
1557	Defensin_ propep	Defensin propeptide	9.8	-11.4	1	1-50
1559	GTP_EFT U	Elongation factor Tu GTP binding domain	1.5e-11	51.8	1	125-348
1559	GTP_EFT U_D3	Elongation factor Tu C-terminal domain	8.1e-07	33.1	1	451-541
1559	GTP_EFT U_D2	Elongation factor Tu domain 2	1e-06	35.8	1	363-446

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1559	ATP-bind	Conserved hypothetical ATP binding protein	8.5	-132.9	1	126-312
1559	dynam	Dynamins family	8.7	-85.0	1	110-278
1567	carb_anhyd	Eukaryotic-type carbonic anhydrase	4.6e-170	578.4	1	5-241
1568	An_peroxidase	Animal haem peroxidase	8.1e-164	557.6	1	144-683
1568	DUF37	Domain of unknown function DUF37	6.5	-36.0	1	462-518
1569	DAO	FAD dependent oxidoreductase	0.055	-90.8	1	49-381
1571	CH	Calponin homology (CH) domain	1.4e-25	98.4	1	126-233
1573	NUDIX	MutT-like domain	5.5e-12	53.3	1	96-221
1574	HECT	HECT-domain (ubiquitin-transferase)	4.3e-16	66.9	1	281-573
1575	ig	Immunoglobulin domain	2.8	11.5	1	122-187
1577	7tm_1	7 transmembrane receptor	3.4	-113.2	1	42-246
1577	Bac_rhodopsin	Bacteriorhodopsin	9.6	-139.9	1	111-313
1578	fn3	Fibronectin type III domain	0.21	11.2	1	121-211
1579	Acetyltransferase	Acetyltransferase (GNAT) family	2.4e-14	61.1	1	548-624
1580	Acetyltransferase	Acetyltransferase (GNAT) family	2.4e-14	61.1	1	653-729
1580	Acetyltransferase	Acetyltransferase (GNAT) family	2.4e-14	61.1	1	653-729
1582	GIDA	Glucose inhibited division protein A	0.0017	-414.5	1	68-196
1583	efhand	EF hand	1.2	14.9	1	23-51
1585	Dynein_h_cavy	Dynein heavy chain	3.6e-18	-92.1	1	1-363
1587	Sm	Sm protein	2e-07	38.1	1	43-124
1588	PDZ	PDZ domain (Also known as DHR or GLGF)	3.1e-15	64.1	1	3-83
1590	MAP1_L_C3	Microtubule associated protein 1A/1B, light	0.04	-35.1	1	99-187
1591	Syntaxin	Syntaxin	2.3e-09	38.1	1	1-266
1591	synaptobrevin	Synaptobrevin	5.8	-14.5	1	184-272
1591	DUF148	Domain of unknown function DUF148	7.7	-38.1	1	17-129
1592	laminin_E_GF	Laminin EGF-like (Domains III and V)	1.2	-4.2	1	153-196
1592	EGF	EGF-like domain	2.4	10.9	3	140-177;284-313;351-380
1592	metalthio	Metallathionein	4.8	-9.3	1	288-348
1593	DnaJ	DnaJ domain	3.4e-40	146.9	1	3-69
1594	HMG_box	HMG (high mobility group) box	2.6e-27	104.1	1	346-414
1598	HMG_box	HMG (high mobility group) box	3.4e-30	113.8	1	45-113
1600	CUB	CUB domain	6.9	-43.0	1	224-313
1601	DUF6	Integral membrane protein DUF6	1.1e-12	55.6	2	113-238;266-390
1601	sugar_transporter	Sugar (and other) transporter	5.7	-191.4	1	55-405
1601	NADH-ubiquinone-plastoquinone (complex I)	NADH-ubiquinone/plastoquinone (complex I)	6	-167.3	1	131-389
1602	FF	FF domain	2.3e-33	124.3	5	272-321;339-388;406-

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						461:486-541:622-673
1602	WW	WW domain	4.3e-20	80.2	2	88-117:129-158
1603	FF	FF domain	2.3e-33	124.3	5	235-284:302-351:369-424:449-504:585-636
1603	WW	WW domain	1.1e-09	45.6	1	92-121
1605	AT_hook	AT hook motif	1.9	11.9	1	360-372
1606	aminotran 5	Aminotransferase class-V	1.3e-127	437.3	1	37-377
1607	aminotran 5	Aminotransferase class-V	3.5e-94	326.3	1	37-331
1611	Granin	Granin (chromogranin or secretogranin)	6.6	-185.2	1	125-609
1612	PHD	PHD-finger	0.59	-7.8	1	551-610
1613	Branch	Core-2/1-Branching enzyme	1e-77	271.6	1	46-313
1614	mbt	mbt repeat	3.2e-101	349.7	4	78-153:192-265:304-381:412-486
1614	SAM_PN T	Sterile alpha motif (SAM)/Pointed domain	0.0021	6.9	1	809-888
1614	SAM	SAM domain (Sterile alpha motif)	0.023	20.6	1	822-883
1615	UPF0103	Protein of unknown function DUF52	4.7e-64	226.2	1	9-270
1616	C2	C2 domain	6.8e-36	132.7	2	606-695:755-842
1617	C2	C2 domain	7.8e-35	129.2	2	87-176:236-323
1618	C2	C2 domain	0.16	2.4	1	265-346
1619	7tm_1	7 transmembrane receptor (rhodopsin family)	3.9e-20	80.3	1	217-427
1620	K_tetra	K ⁺ channel tetramerisation domain	1.1e-25	98.7	1	3-101
1620	BTB	BTB/POZ domain	9	-22.4	1	21-104
1623	cyclin	Cyclin, N-terminal domain	0.057	-1.4	1	46-149
1624	zf-C2H2	Zinc finger, C2H2 type	8.9e-19	75.8	4	34-57:71-93:112-134:143-165
1627	PDZ	PDZ domain (Also known as DHR or GLGF)	4.2e-16	66.9	1	66-149
1628	MttA_Hcf 106	mttA/Hcf106 family	2.5	-14.8	1	54-107
1628	Tropomyo sin	Tropomyosin	4	-114.1	1	74-272
1628	Syntaxin	Syntaxin	5.7	-103.5	1	82-402
1628	PI3_P14 kinase	Phosphatidylinositol 3- and 4-kinase	8.5	-118.0	1	125-342
1628	HlyD	HlyD family secretion protein	9.4	-64.2	1	129-400
1628	UPF0089	Uncharacterised protein family	10	-132.7	1	150-338
1629	rrm	RNA recognition motif	1.7e-47	171.2	2	72-

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						142:156-226
1632	Ribosomal L10	Ribosomal protein L10	4.2	-31.6	1	70-173
1632	Ribosomal L10	Ribosomal protein L10	4.2	-31.6	1	70-173
1635	death	Death domain	1.5	2.0	1	261-348
1636	death	Death domain	1.5	2.0	1	73-160
1637	Gelsolin	Gelsolin repeat	4.1e-92	319.4	6	27-76:148-188:265-307:398-451:523-564:626-668
1639	TBC	TBC domain	2e-08	15.3	1	98-293
1640	TBC	TBC domain	2e-07	1.3	1	98-297
1641	homeobox	Homeobox domain	0.0097	7.0	1	83-135
1646	transmembrane4	Tetraspanin family	1.6e-75	264.3	1	18-264
1652	LacY_sym	LacY proton/sugar symporter	4.9	-335.5	1	66-299
1652	oxidored_q1	NADH-Ubiquinone/plastoquinone (complex I)	7.6	-169.3	1	53-279
1653	Colipase_C	Colipase, C-terminal domain	4.3e-24	93.5	1	21-65
1654	SH2	SH2 domain	0.0091	3.5	1	1-83
1658	SUI1	Translation initiation factor SUI1	4.4e-46	166.5	1	50-149
1661	jmjC	jmjC domain	0.00052	13.9	1	308-415
1663	SH3	SH3 domain	0.011	9.5	1	332-388
1664	UBA	UBA/TS-N domain	6.7e-06	33.0	1	194-233
1664	TUDOR	Tudor domain	0.2	-0.7	1	506-627
1665	Ribosomal S21	Ribosomal protein S21	0.0039	11.7	1	10-62
1666	transmembrane4	Tetraspanin family	2.2e-71	250.6	1	23-264
1669	RuBisCO_small	Ribulose biphosphate carboxylase, small	2.8	-52.1	1	354-430
1671	LRR	Leucine Rich Repeat	2e-50	180.9	12	29-47:48-71:72-95:96-118:119-142:143-166:167-189:190-213:214-235:236-259:260-283:284-307
1671	7tm_1	7 transmembrane receptor	0.0032	-43.2	1	439-688
1671	Cytidyltransferase	Phosphatidate cytidyltransferase	7.1	-89.9	1	520-617
1671	oxidored_q1	NADH-Ubiquinone/plastoquinone	9.7	-171.5	1	475-685
1671	MerC	MerC mercury resistance protein	9.8	-87.5	1	534-632
1672	UCH-2	Ubiquitin carboxyl-terminal hydrolase family	4e-28	106.9	1	292-364
1672	UCH-1	Ubiquitin carboxyl-terminal hydrolases	1.9e-14	61.4	1	35-66

SEQ ID	Model	Description	E-value	Score	Repeats	Position
		family				
1673	F-box	F-box domain	0.00071	26.3	1	87-134
1674	Lamp	Lysosome-associated membrane glycoprotein	5.7	-191.1	1	351-653
1675	FGGY	FGGY family of carbohydrate kinases, N-term	2.7e-45	163.9	1	20-282
1675	FGGY_C	FGGY family of carbohydrate kinases, C-term	5.5e-23	89.8	1	285-491
1676	Keratin_B2	Keratin, high sulfur B2 protein	7.4	-81.9	1	24-201
1678	S_100	S-100/ICaBP type calcium binding domain	8.3	-9.6	1	909-941
1680	pyr_redox	Pyridine nucleotide-disulphide oxidoreductase	7.4e-24	92.7	1	12-369
1681	WD40	WD domain, G-beta repeat	9.3e-14	59.1	6	305-343:352-389:395-432:440-476:487-524:583-621
1683	Ribosomal L44	Ribosomal protein L44	1e-38	142.1	1	17-95
1686	WD40	WD domain, G-beta repeat	8.9e-14	59.2	2	8-43:50-86
1688	FlIE	Flagellar hook-basal body complex protein Fl	5.1	-27.5	1	673-763
1690	dCMP_cyt deam	Cytidine and deoxycytidylate deaminase	2e-13	58.0	1	12-100
1691	G-gamma	GGL domain	5.1	-8.6	1	712-760
1692	cpn60_TC P1	TCP-1/cpn60 chaperonin family	0.012	-260.5	1	32-187
1693	Glycophorin A	Glycophorin A	4.4	-43.1	1	16-149
1696	zf-C2H2	Zinc finger, C2H2 type	4.5e-15	63.5	6	92-115:120-143:174-198:210-233:329-353:363-386
1698	LRR	Leucine Rich Repeat	0.44	17.0	4	37-58:59-80:81-102:103-125
1699	VHS	VHS domain	9.5e-60	211.9	1	5-146
1700	rrm	RNA recognition motif	3.6e-23	90.4	3	128-203:332-402:413-480
1701	ank	Ankyrin repeat	3.8e-101	349.4	12	12-44:45-77:79-111:112-144:145-177:179-211:212-244:245-277:278-310:312-